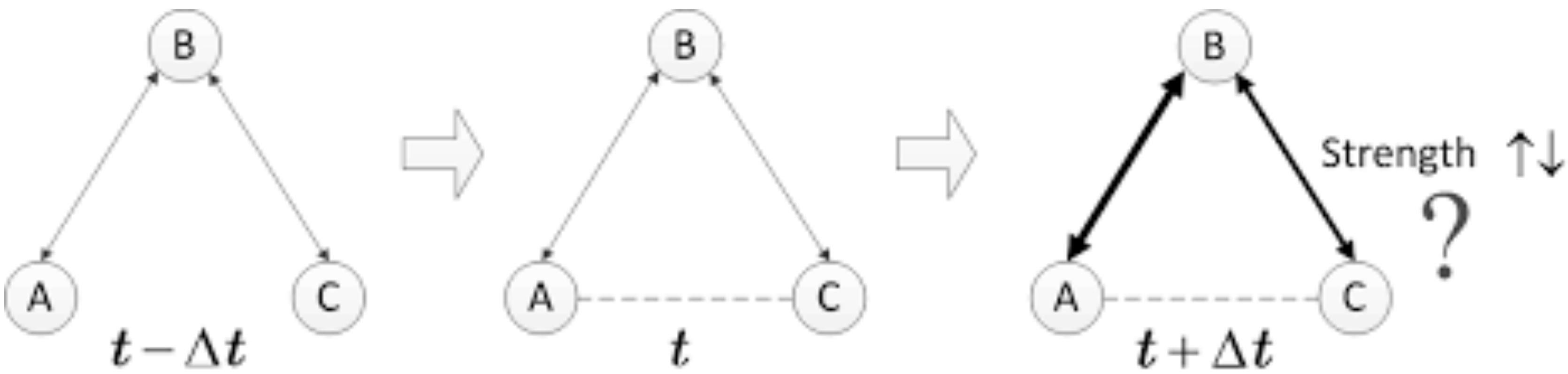
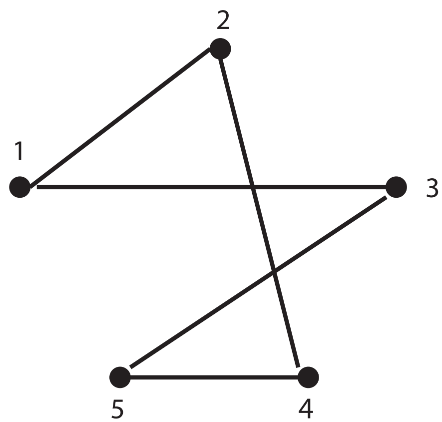
Network Concepts from Sociology

Triadic Closure and Clustering Coefficient



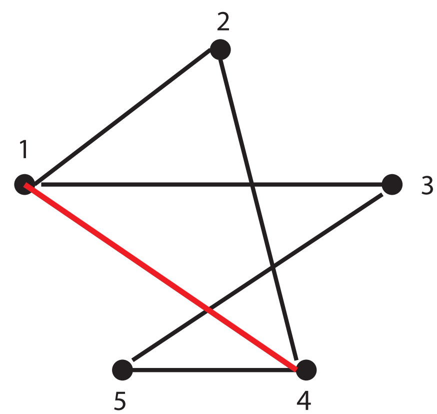
Triadic Closure

Many network concepts have come from sociology, studying interactions of groups of people. Ramifications here are huge for financial interests, the spread of information (and misinformation), and the spread of epidemics.



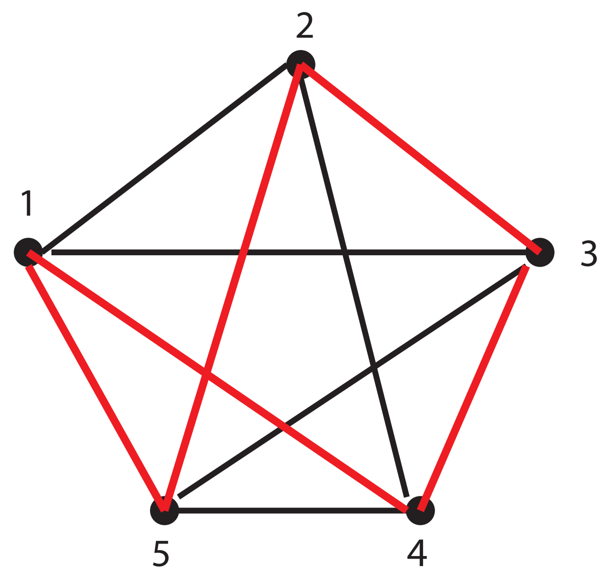
Consider this friendship graph. Person 1 and person 4 have a friend in common, person 2.

So it is likely that 1 and 4 will themselves become friends. This is an example of triadic closure.



Triadic Closure

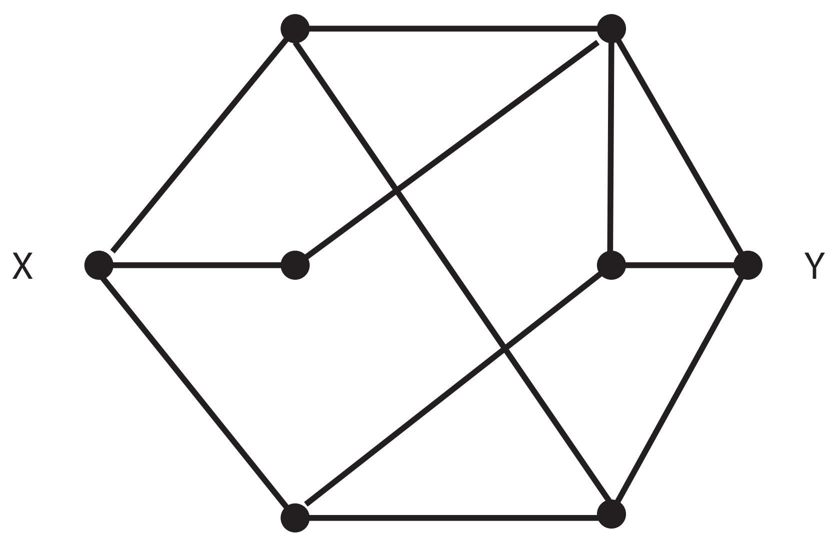
As a result of this triadic closure, a triangle has been formed in the network. Additional triadic closure will result, over time, in more triangles.



Clustering Coefficient (C)

This is a measure that attempts to capture the degree of triadic closure in a network. The clustering coefficient of a node A is defined as the probability that two randomly selected friends of A are friends with each other. If there are q friends, then there are $\binom{q}{2}$ friend pairs.

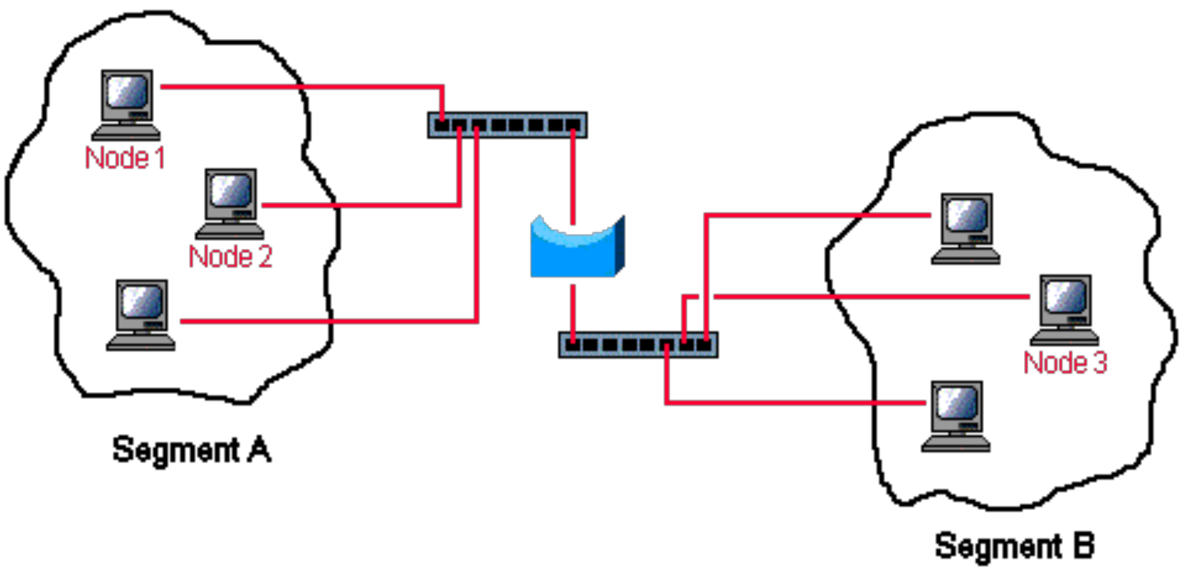
That is, it is the fraction of pairs of A's friends that are connected to each other by edges.



$$C_X = \frac{0}{3} = 0$$

$$C_Y = \frac{1}{3}$$

Bridges and Ties



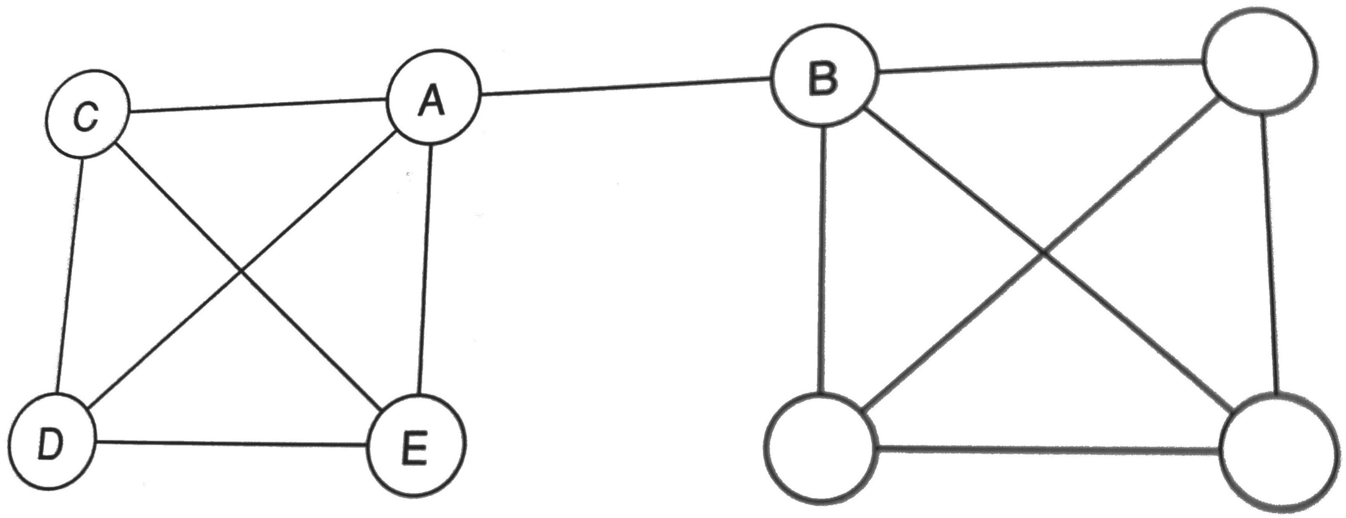
Bridges and Local Bridges

In a study by Mark Granovetter for his PhD, he conducted interviews with a large number of people and determined that many found jobs through acquaintances. That's not surprising. What is surprising is that this number was greater than the number of people who found jobs through close friends.

WHY?

Bridges Between Cliques

Easley and
Kleinberg 2010

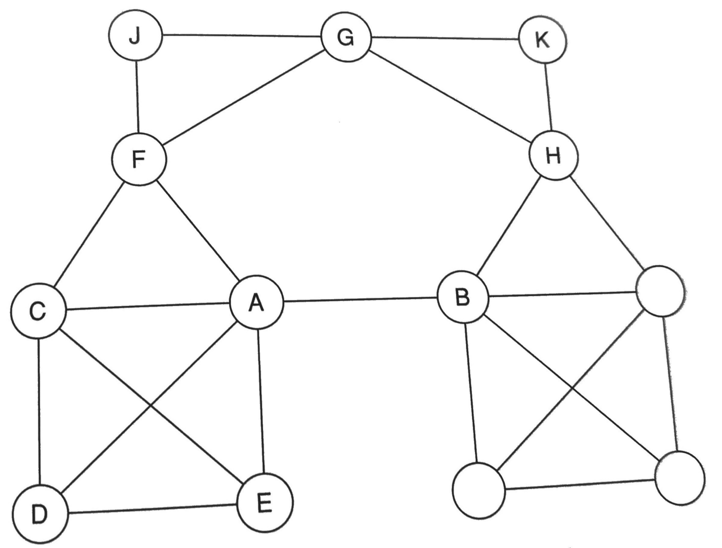


The edge joining A and B in this network is special. If it is removed, then A and B lie in two different components of the graph. Such an edge is called a **bridge**, since it bridges components of the graph.

In terms of a social network, a bridge would be a friendship that couples two different clusters of friends. Each such cluster is called a clique.

Local Bridges are More Common

Easley and
Kleinberg 2010



Bridges are very rare in social networks, and usually disappear when larger networks of individuals are considered. Nodes A and B might think that their friendship is all that couples their respective cliques of friends, but actually there are other connections they might not know about. We should therefore define a less restrictive type of bridge.

An edge joining two nodes A and B is a local bridge if its endpoints have no friends in common – in other words, deleting the edge would increase the distance between A and B to a value greater than 2. This distance is called the span of a local bridge between A and B.

Local Bridges and Triadic Closure

If the A-B edge is a local bridge, can it also be part of a triangle in the graph?

No, because if it is part of a triangle, then if removed, the distance between A and B would be 2 (not greater than 2).

Weak and Strong Ties

What's the difference between a friend and an acquaintance? A friend is someone with strong ties, while an acquaintance has weaker ties.

We can weight the edges as W=weak and S=Strong.

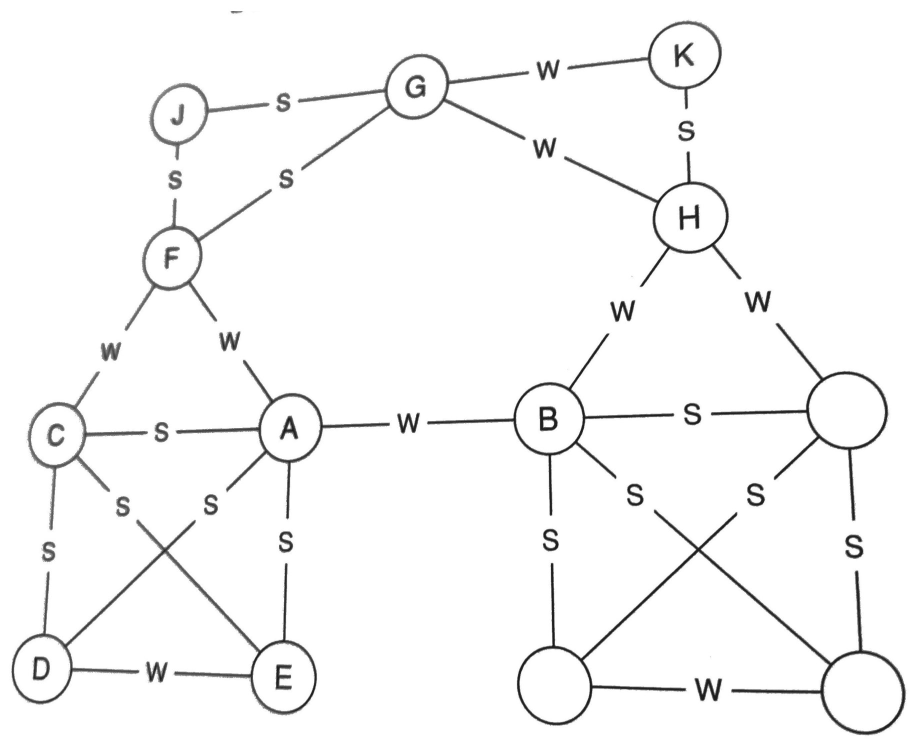
In the last graph, do you think the A-B edge would be weak or strong?

Probably weak, since A and B have no common friends.

In most cases, local bridges are weak ties.

Weak and Strong Ties

Easley and
Kleinberg 2010



Acquaintances are better at recommending jobs since they are in touch with an entirely different clique of people who might have heard something about the job. In contrast, your friends are likely to have heard the same things you have, so will provide no new information.

Generalizing the Notions of Tie Strength and Local Bridges

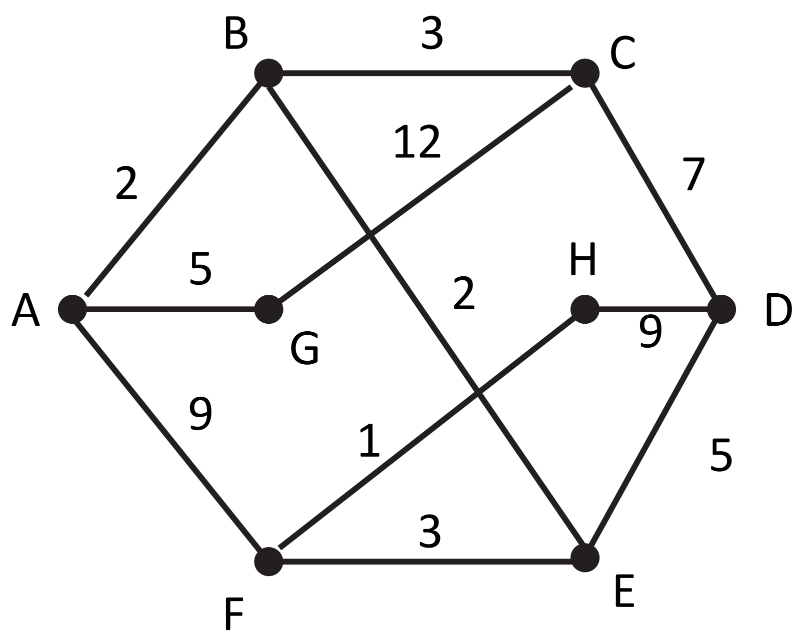
Rather than assigning edges W and S strengths, one could use a real number. This might correspond to the number of minutes of direct or cell phone conversation per month, for example.

Since there are so few local bridges in most social networks, it makes sense to soften the definition to include "almost" local bridges. To this end, define the neighborhood overlap of an A-B edge as:

$$\boxed{\text{Overlap} = \frac{\text{number of nodes who are neighbors of both A and B}}{\text{number of nodes who are neighbors of at least one of A or B}}}$$

The neighborhood overlap is 0 when an edge is a true local bridge. If the overlap is near 0, then the edge is almost a local bridge.

Relationship Between Tie Strength and Neighborhood Overlap

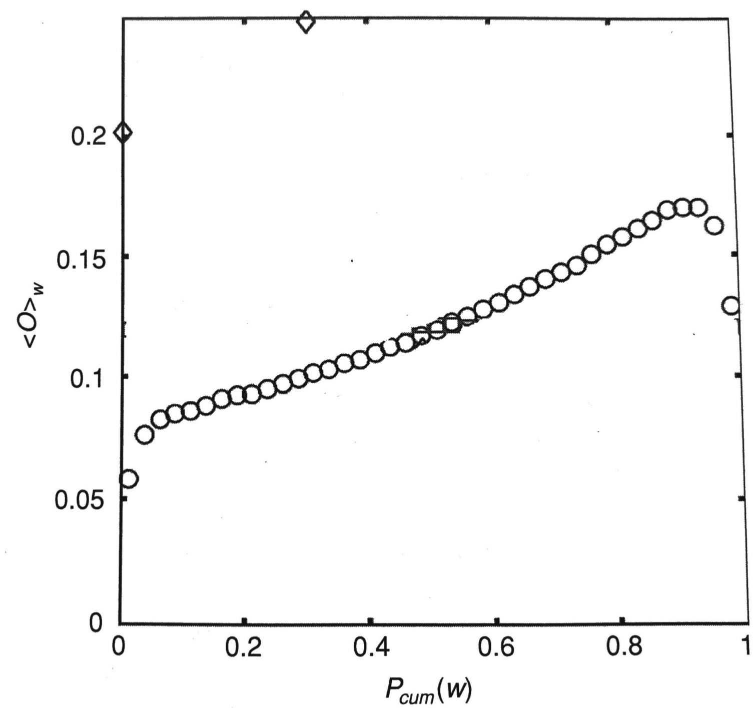


Sort by tie strength:

F-H
A-B
B-E
E-F
B-C
etcetera

Relationship Between Tie Strength and Neighborhood Overlap

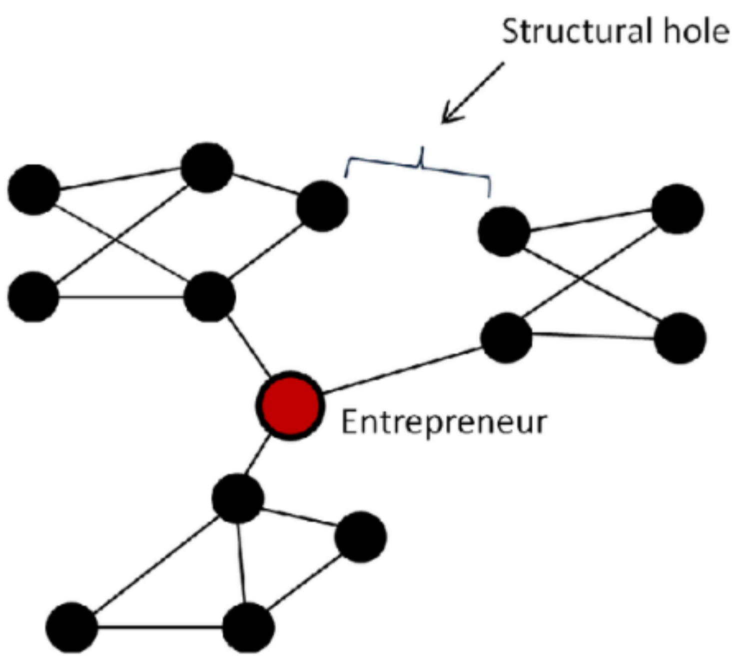
Easley and Kleinberg 2010



Example of real social network using mobile phone data for tie strength.
X-axis: ordering of edges by tie strength
Y-axis: neighborhood overlap

Edges with stronger tie strength have more neighborhood overlap

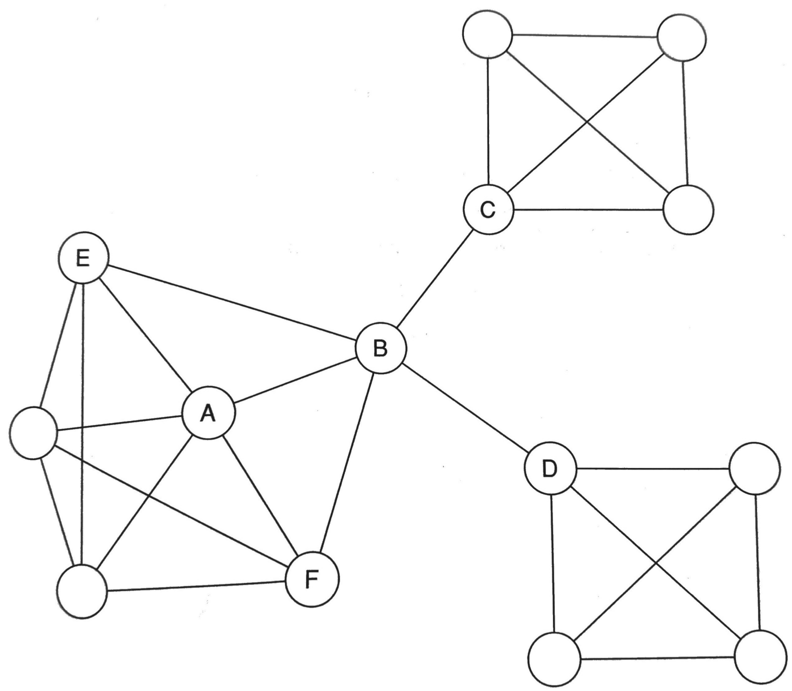
Embeddedness and Structural Holes



Embeddedness

The embeddedness of an edge is the number of common neighbors shared by the two endpoints. This is the numerator in the neighborhood overlap.

Easley and Kleinberg 2010

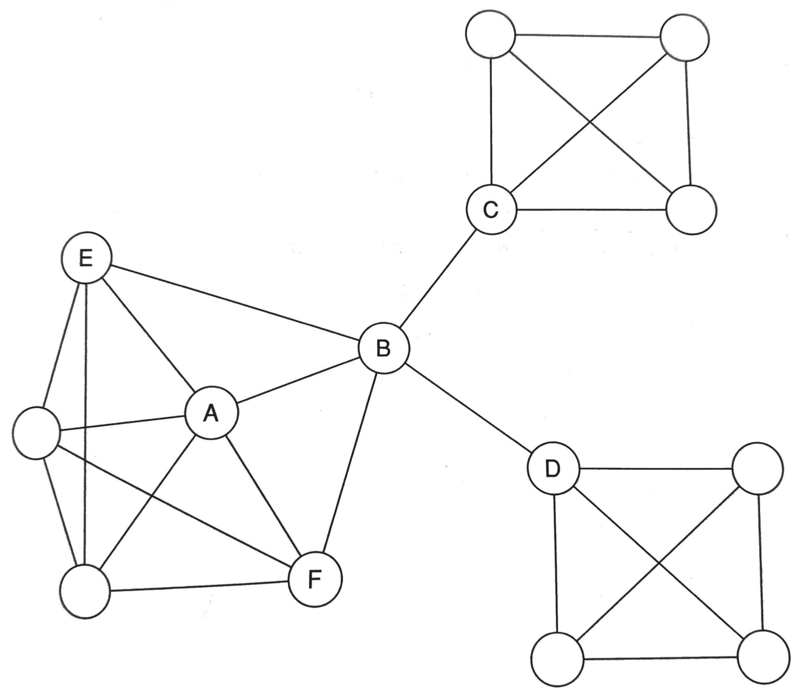


In this example, A-B has embeddedness 2, since A and B have two common neighbors, E and F. Edge B-C is a local bridge, with embeddedness of 0.

Embeddedness

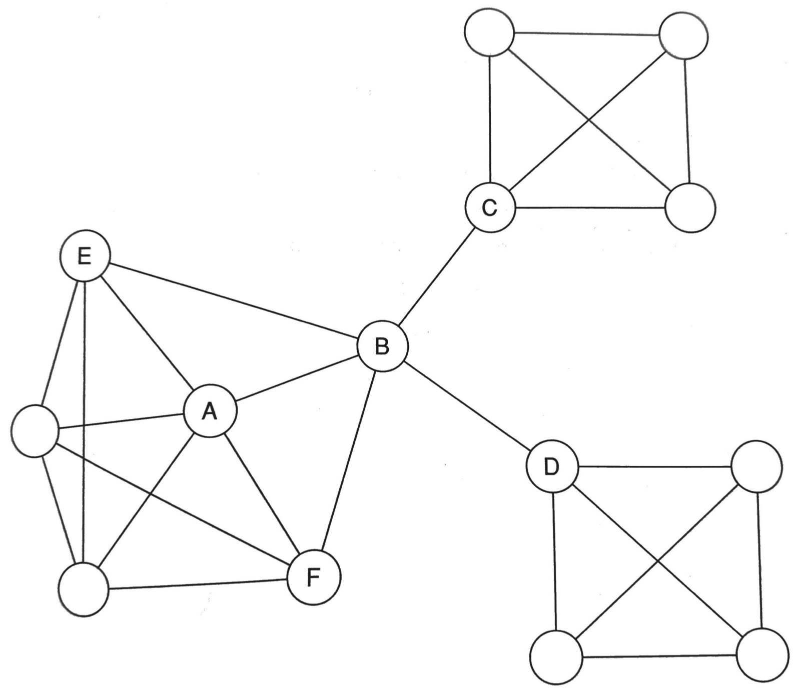
Node A is clearly part of a clique. This is quantified by the fact that each of its edges has high embeddedness. This is in stark contrast to node B, whose edges have lower (or 0) embeddedness.

Easley and Kleinberg 2010



A great deal of sociological research shows that if two individuals are connected by an embedded edge, then it is easier for them to trust each other and any potential transactions they may have together.

Structural Holes

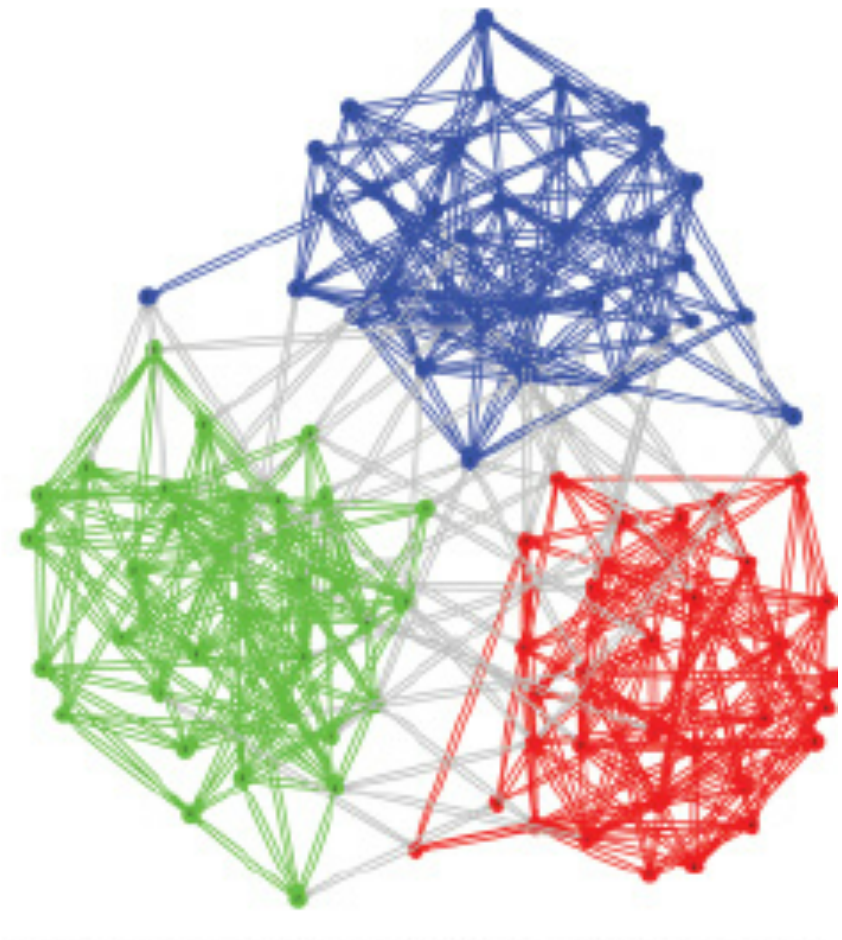


Easley and Kleinberg 2010

There are disconnections or empty spaces between the three clusters of nodes. These are examples of structural holes. Node B fills those holes. Why might that be advantageous for this individual?

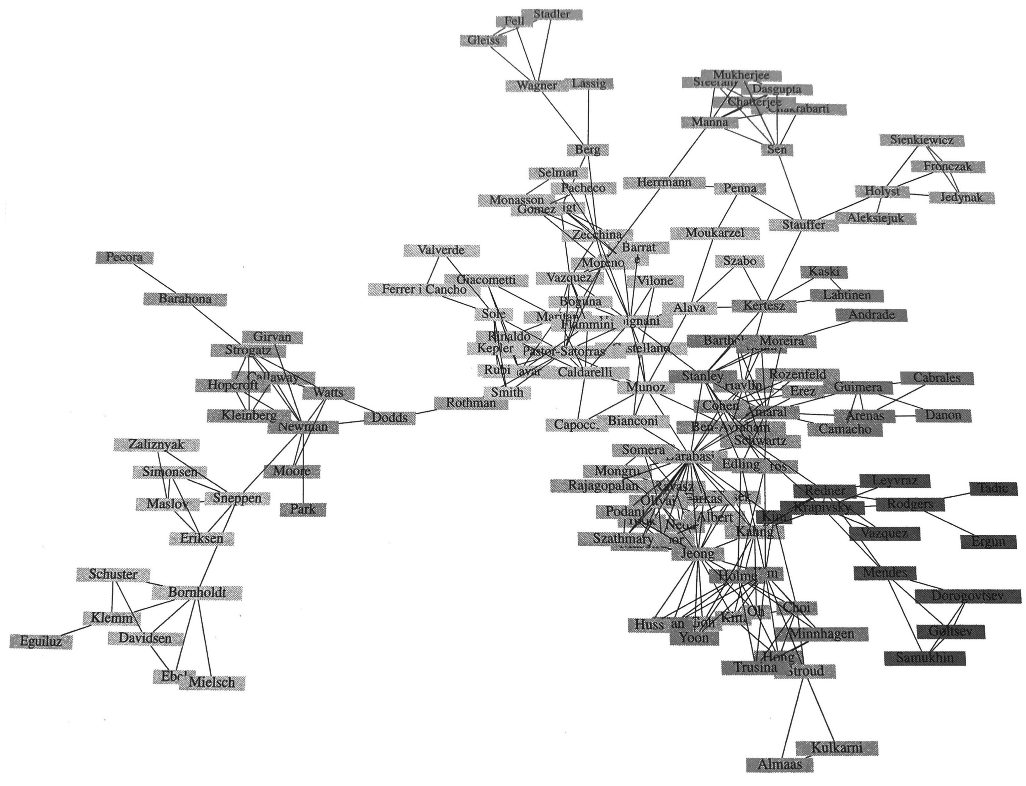
B has access to disparate information that can be combined in novel ways. B is also the gatekeeper between the different groups.

Network Partitioning



Most Networks Contain Highly Connected Clusters

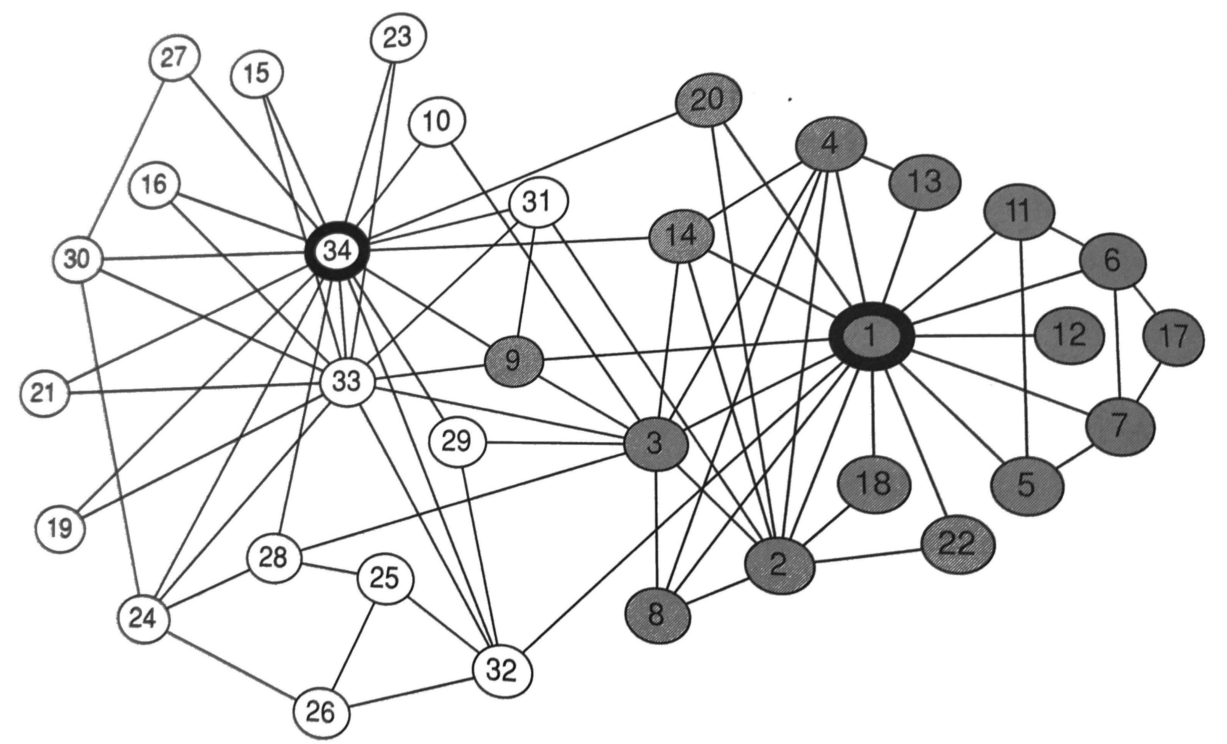
Easley and Kleinberg 2010



Coauthorship network of researchers working on networks.

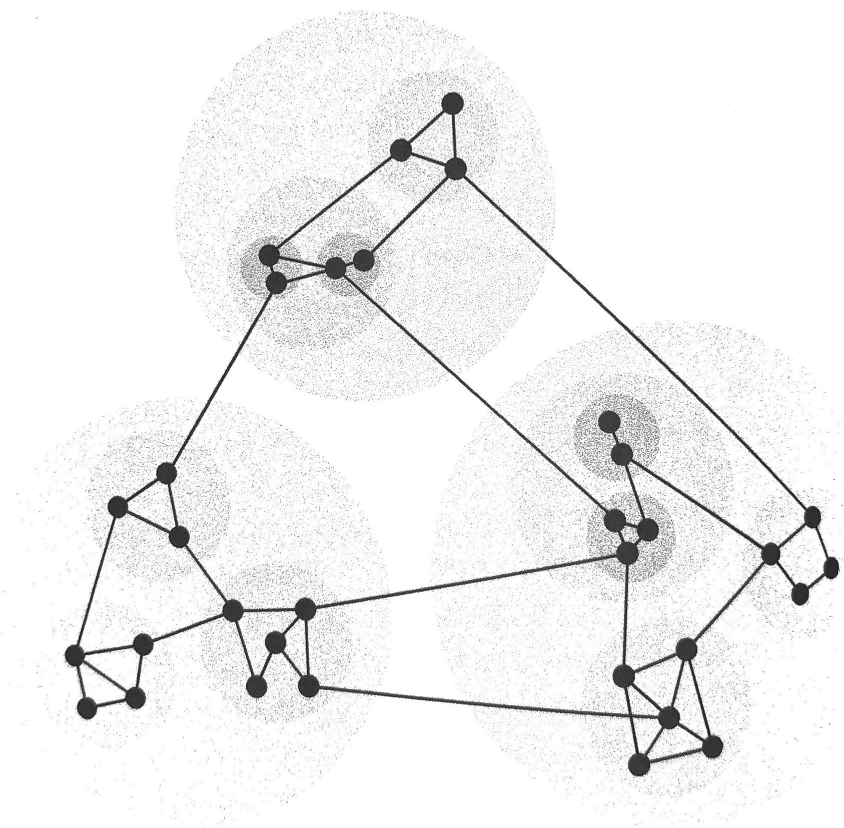
Most Networks Contain Highly Connected Clusters

Easley and
Kleinberg 2010



Members of a karate club that ultimately split into two clubs following a dispute. One group sided with the club president (node 34) and the other with the club instructor (node 1).

Networks Often Have a Hierarchical Structure



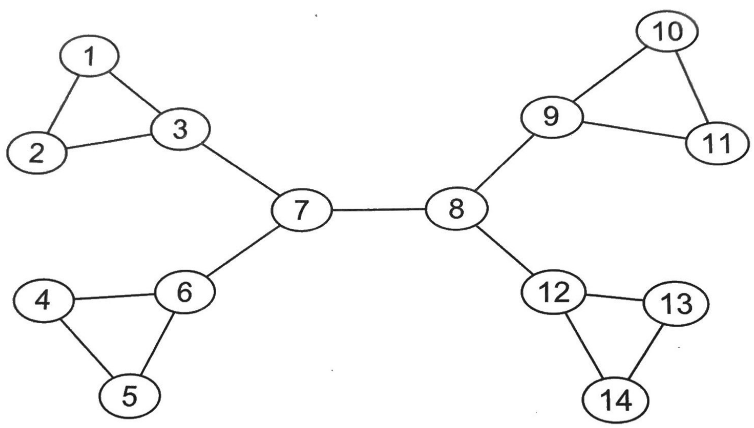
Newman 2018

Clusters can be further subdivided into subclusters, which can be further divided again.

How Can You Identify the Clusters?

Divisive Methods for Network Partitioning

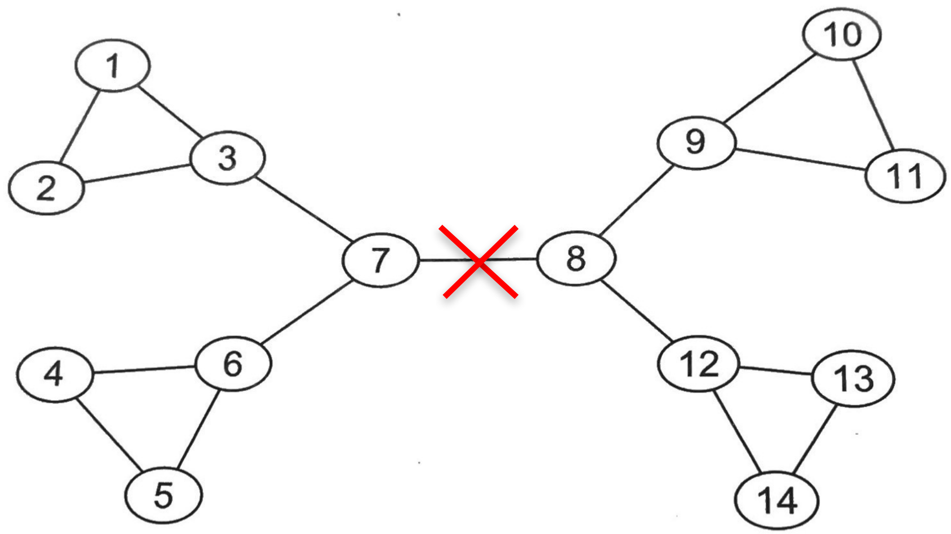
Easley and
Kleinberg 2010



Divisive methods erase edges to divide up the network

Divisive Methods for Network Partitioning

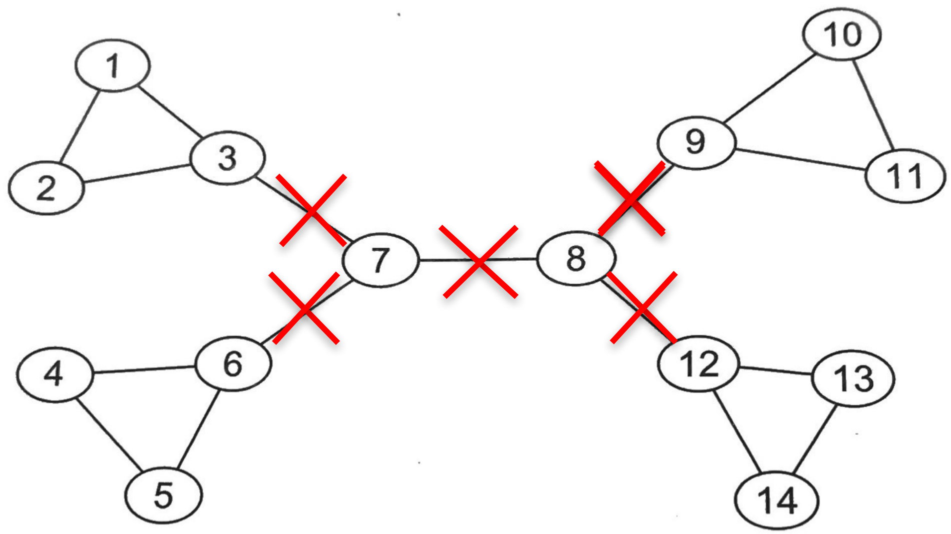
Easley and
Kleinberg 2010



Iteration 1: remove central link

Divisive Methods for Network Partitioning

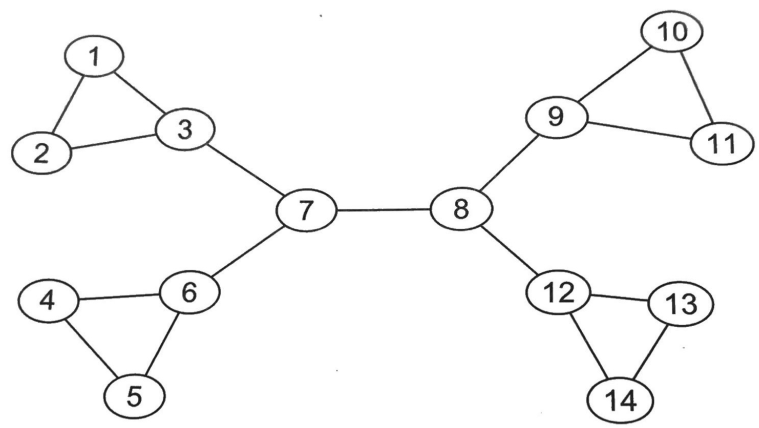
Easley and
Kleinberg 2010



Subsequent iterations: remove links to other clusters

Agglomerative Methods for Network Partitioning

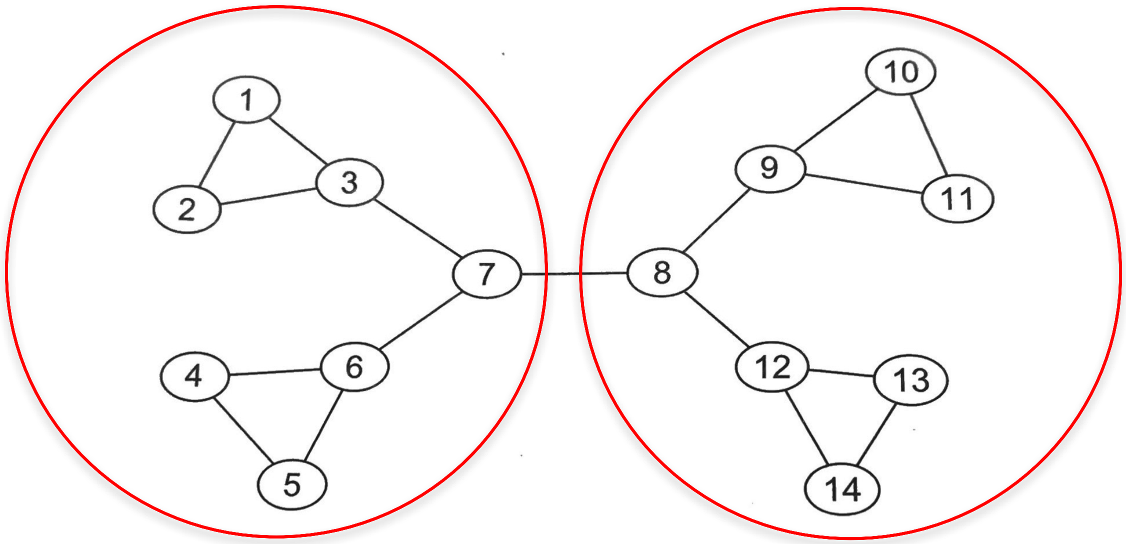
Easley and Kleinberg 2010



Agglomerative methods focus on the most tightly-knit parts of the network, rather than the connections at their boundaries.

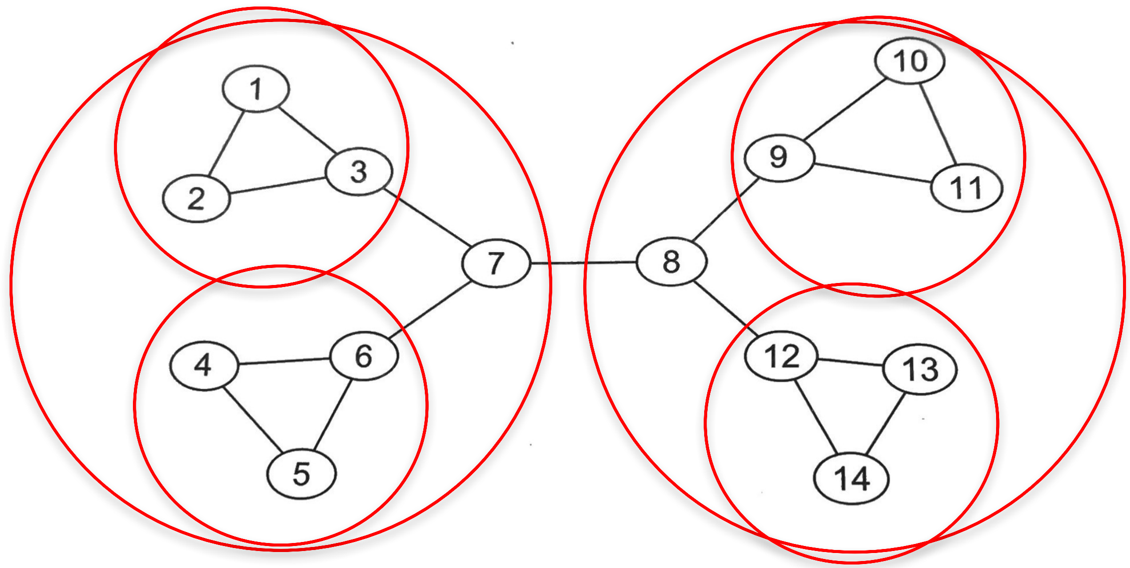
Agglomerative Methods for Network Partitioning

Easley and Kleinberg 2010



Iteration 1: Formally group the biggest clusters

Agglomerative Methods for Network Partitioning



Easley and
Kleinberg 2010

Subsequent iterations: Group smaller clusters
that are contained within the bigger clusters

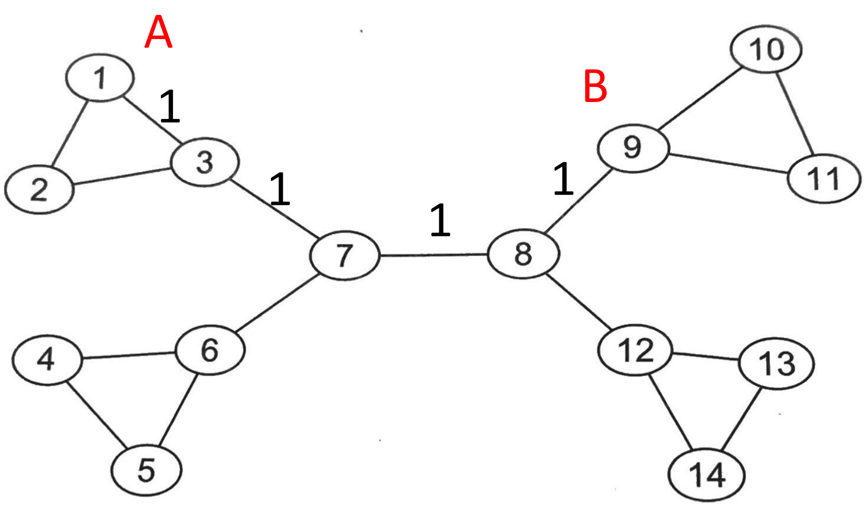
Both types of approaches often produce similar partitions and demonstrate that networks are usually composed of nested subnetworks.

The Girvan-Newman Divisive Method

Basic idea: Iteratively delete the edges containing the most "traffic".

Consider each pair of nodes, A and B. Find the shortest path(s) between them. For each edge on such a path add one unit of flow. If there are two shortest paths, then edges on each gets ½ unit of flow, etc.

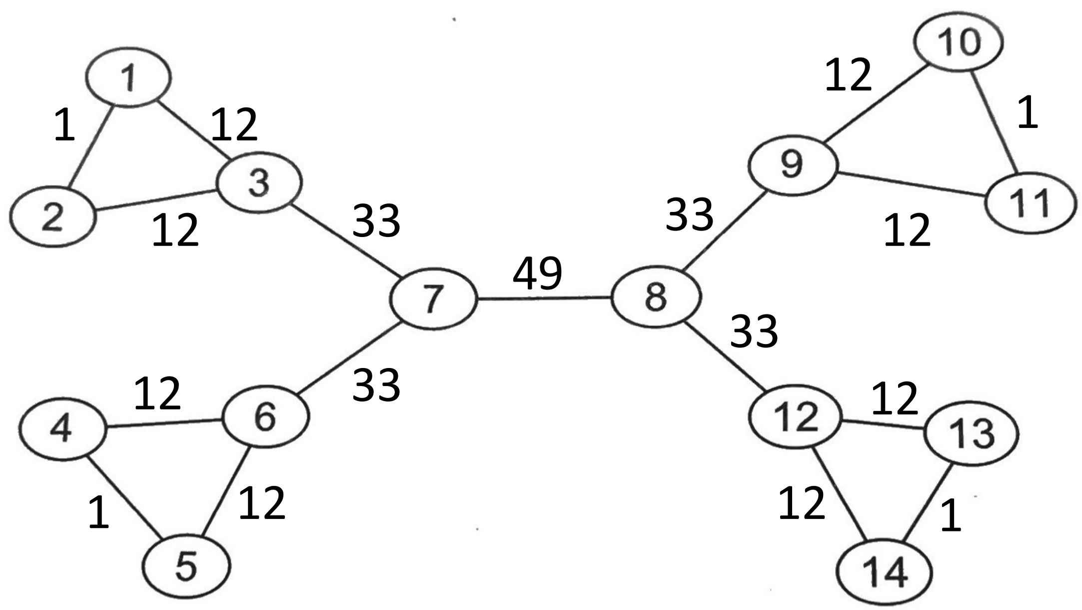
Easley and
Kleinberg 2010



The Girvan-Newman Divisive Method

Sum the flows at each edge for all node pairs. This gives the betweenness of each edge (previously we discussed betweenness of nodes).

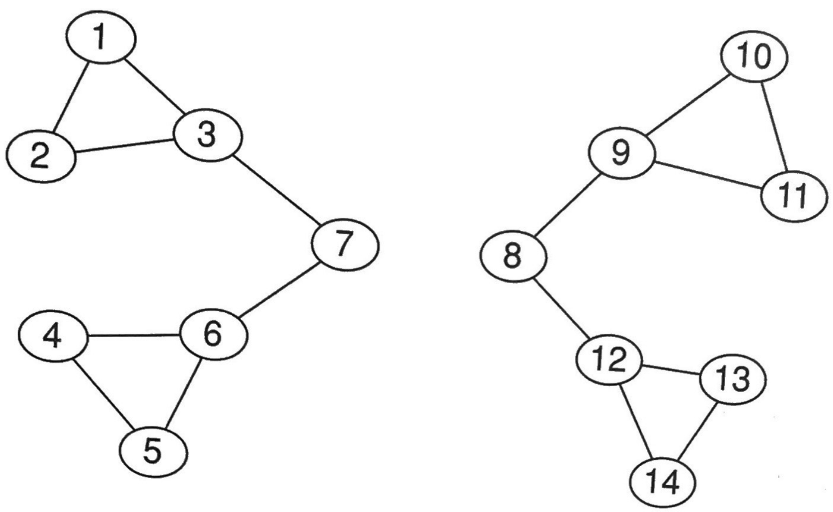
Easley and Kleinberg 2010



The Girvan-Newman Divisive Method

Next, remove the edge with the greatest betweenness (or edges if there is a tie).

Easley and Kleinberg 2010

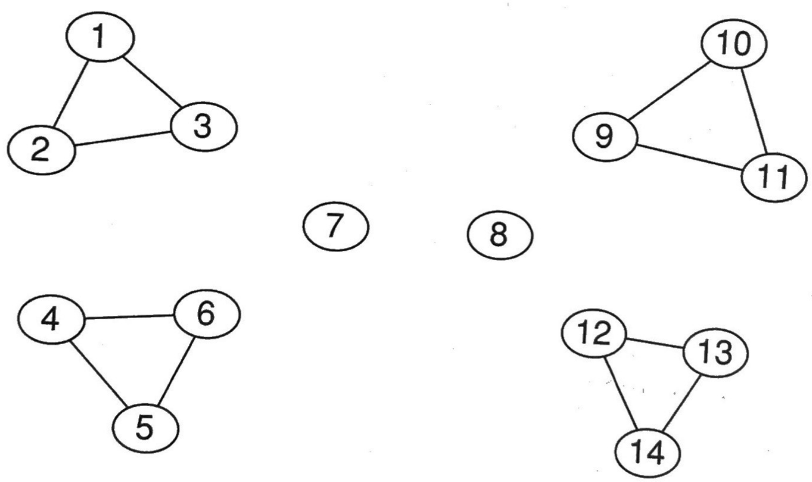


This is the first level of partitioning of the graph.

The Girvan-Newman Divisive Method

Recalculate the betweenness, then iterate again, removing the edge(s) with greatest betweenness.

Easley and Kleinberg 2010

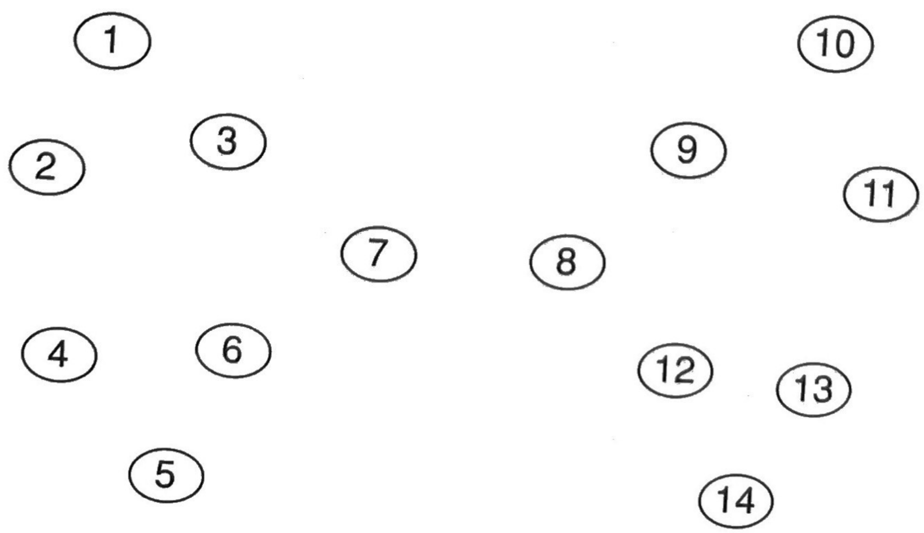


This is the second level of partitioning.

The Girvan-Newman Divisive Method

Repeat until all edges have been removed, each time recalculating the betweenness of all remaining edges.

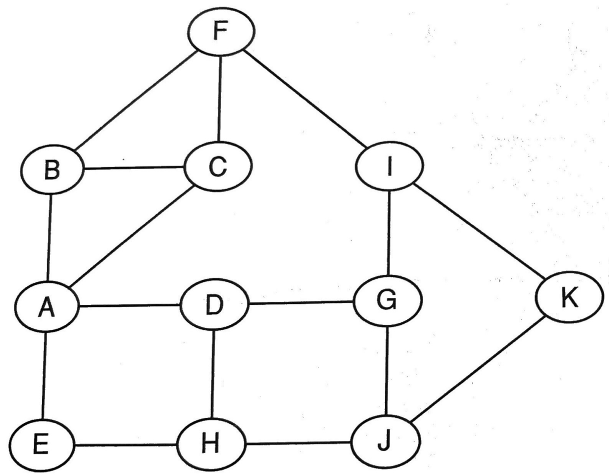
Easley and Kleinberg 2010



Computing Edge Betweenness

Is there a good way to count all the shortest paths between nodes? Yes, there is a clever process based on breadth-first search.

Easley and Kleinberg 2010



Perform a breadth-first search of the graph, starting at A

Determine number of geodesic paths from A to each other node

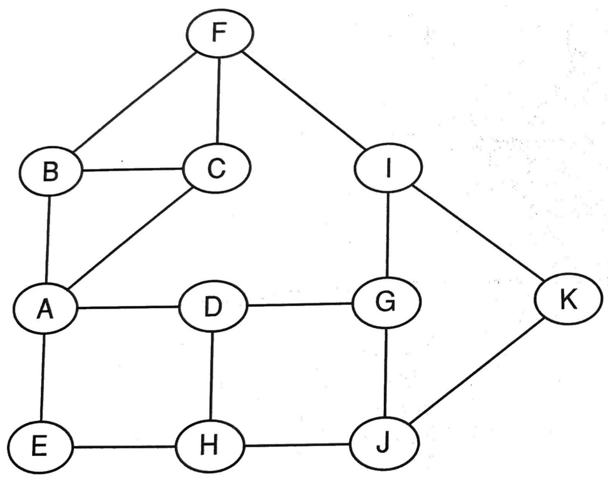
Based on these numbers, determine the amount of flow from A to all other nodes that use each edge

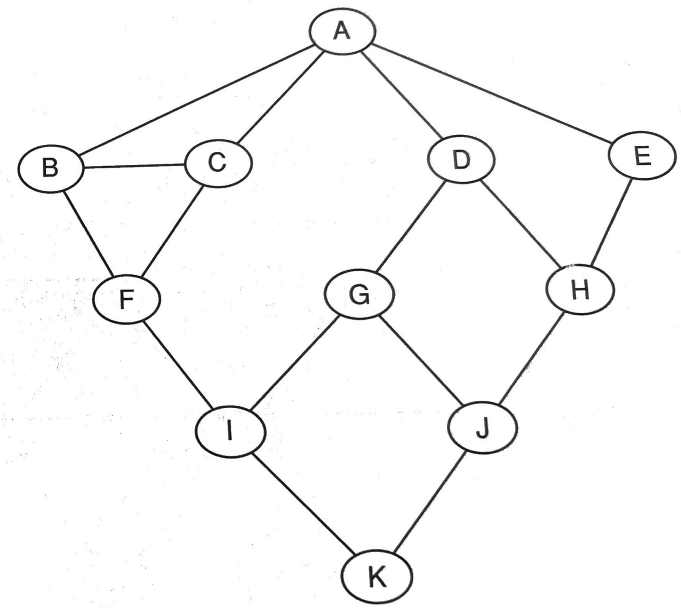
Do this for all other nodes and sum up flows.

Breadth-First Search

Rearrange graph into layers, with A at the top.

Easley and
Kleinberg 2010





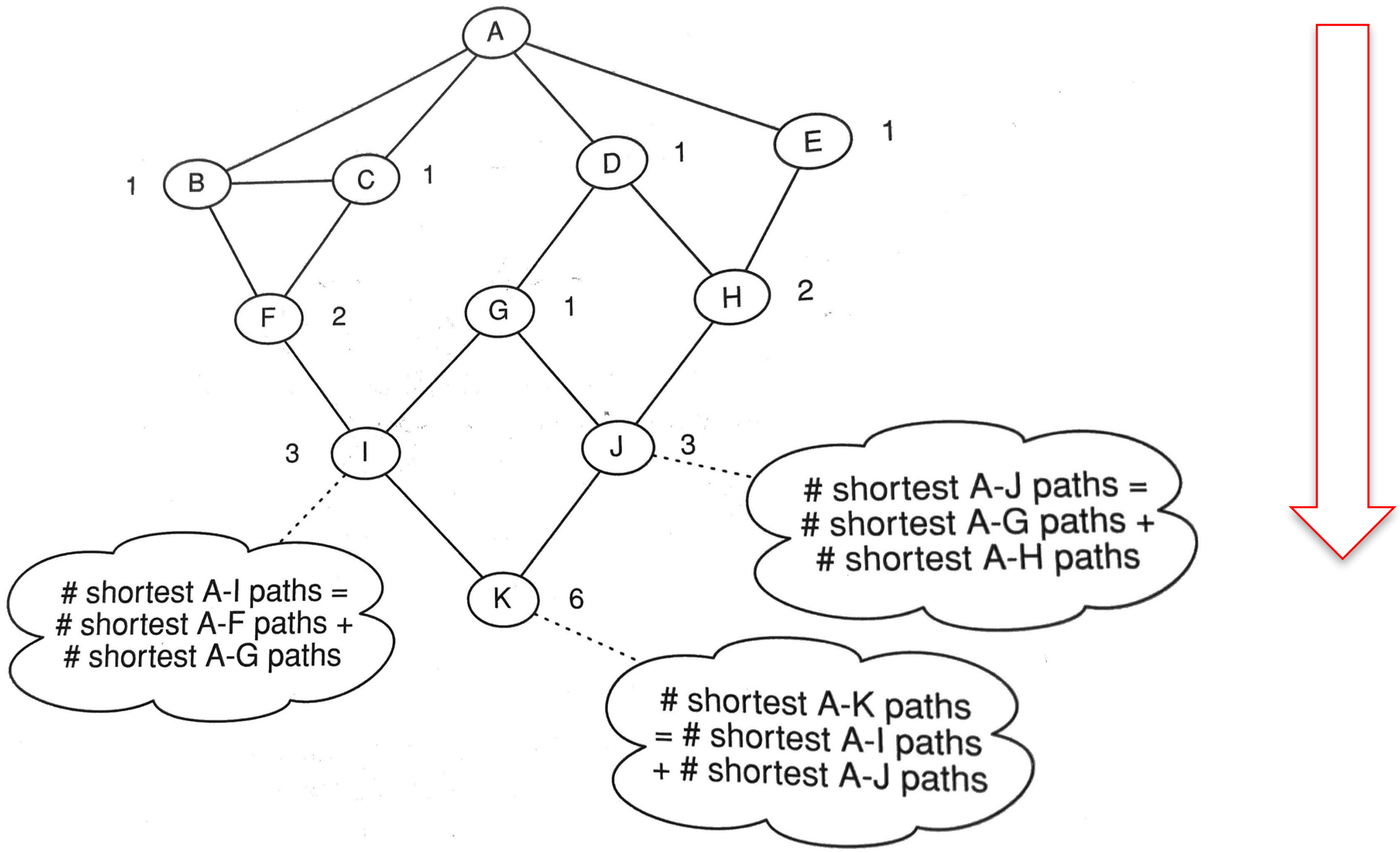
distance=1

distance=2

distance=3

distance=4

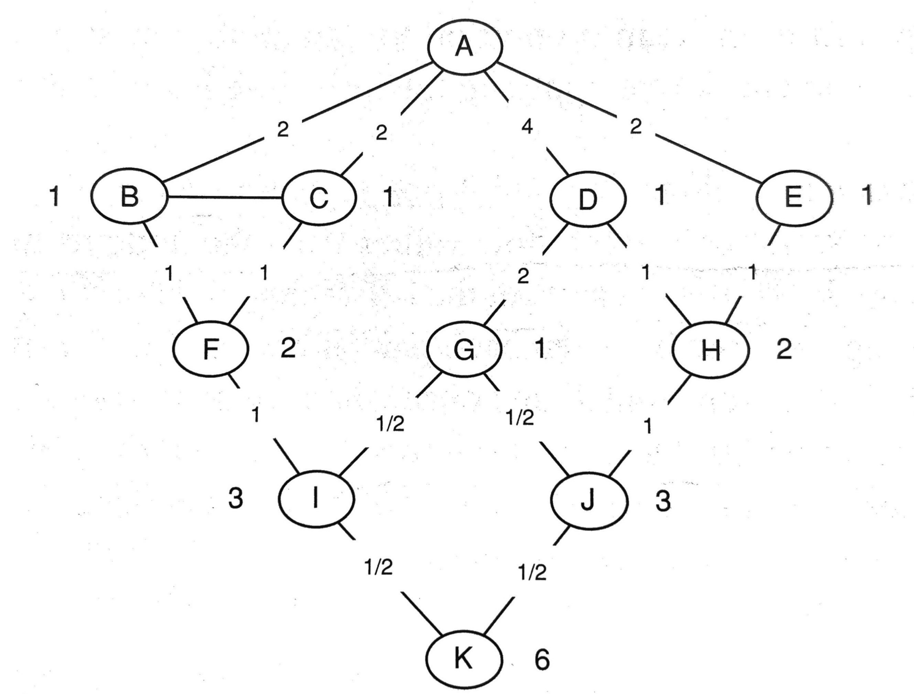
Counting Shortest Paths



Easley and
Kleinberg 2010

Add up number of shortest paths, moving downward through the network.

Determining Flow Values



Easley and
Kleinberg 2010

From bottom, K gets 1 unit of flow, divided equally along I-K and J-K

I gets 1 unit, plus the half unit passing through I to K, so 1 ½ total.
Edge F-I should get twice as much of this than G-I. So F-I gets 1 unit, G-I gets ½.

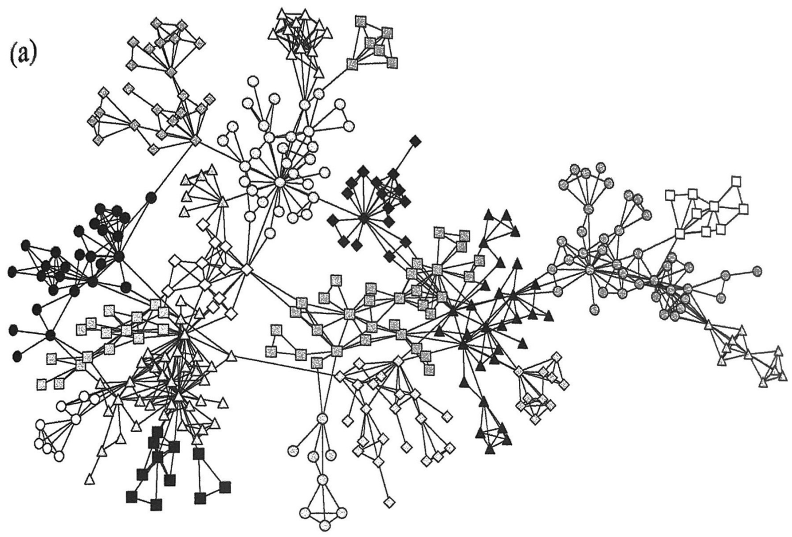
Continue working up through the graph to get all flows associated with node A.

Now repeat, centering on node B. Add flows together. Continue repeating and summing. This gives betweenness values for all edges.

Community Detection

Once a network has been partitioned, what can you do with this information? The example below shows the result of network partitioning.

Newman 2018

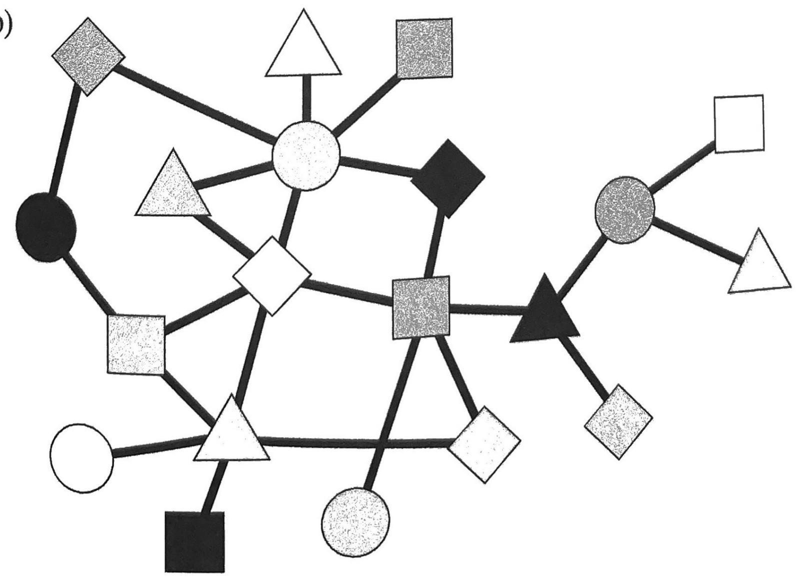


Community Detection

You can now form a new graph, replacing the clusters with single nodes. This is called community detection, and is an example of course graining.

Newman 2018

(b)

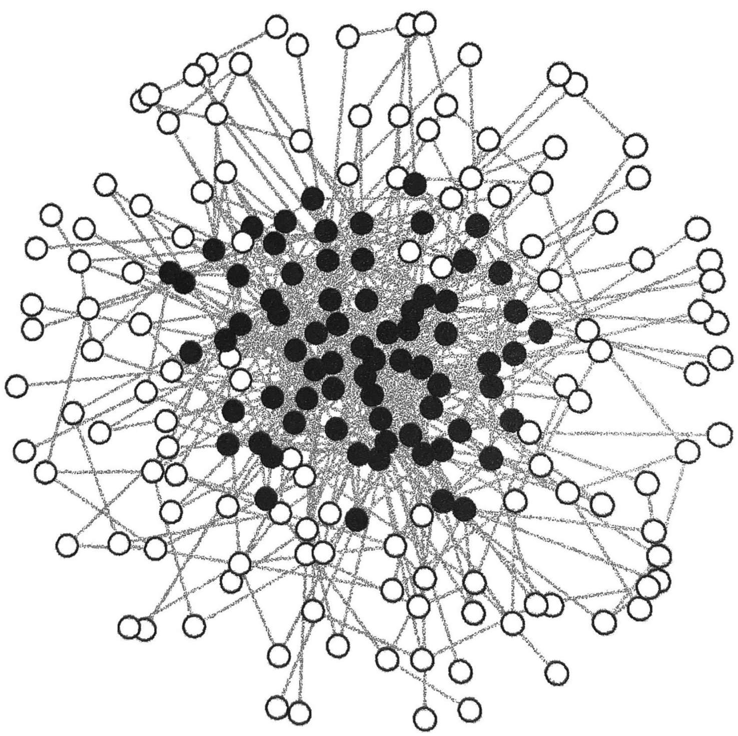


It is much easier to see the community structure of the network in this course-grained figure than in the original network representation, which contained more information than was needed.

Core-Periphery Structure

Many networks have a core of densely-connected nodes with a periphery of nodes with sparser connectivity. An example is the internet.

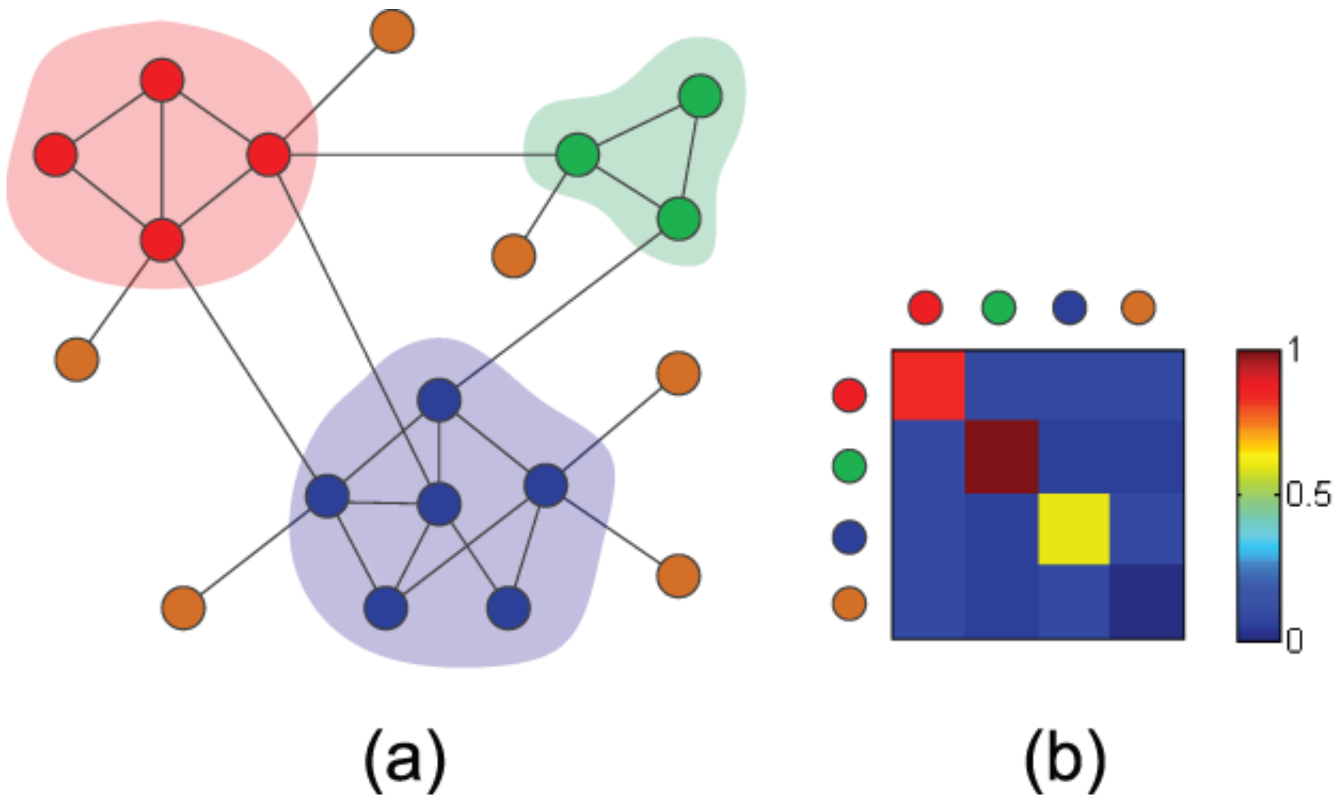
Algorithms exist to distinguish the core from the periphery.



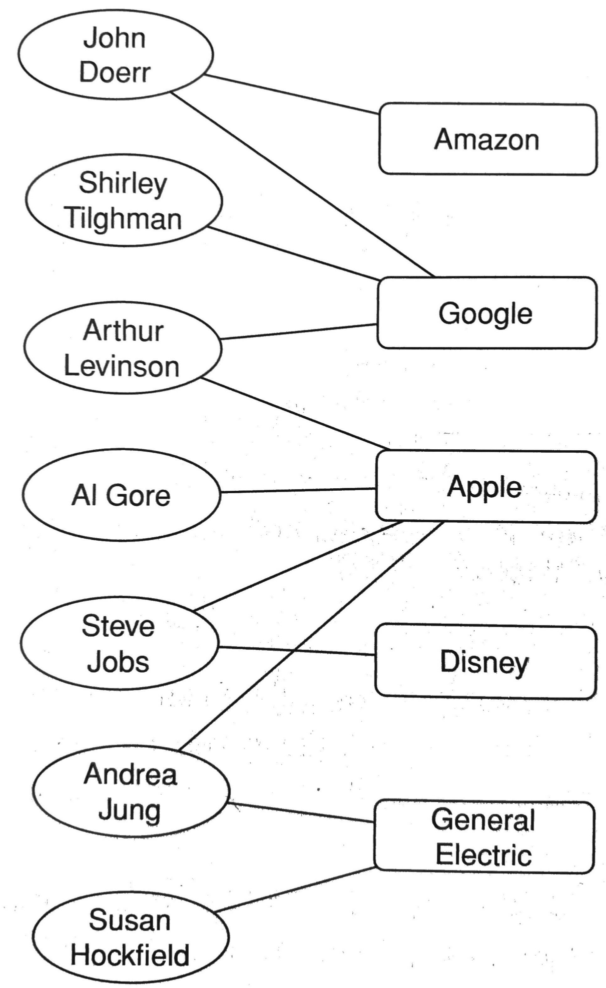
Core nodes are solid, while peripheral nodes are open

Newman 2018

Link Formation in a Social Network



Closure is Influenced by Affiliation



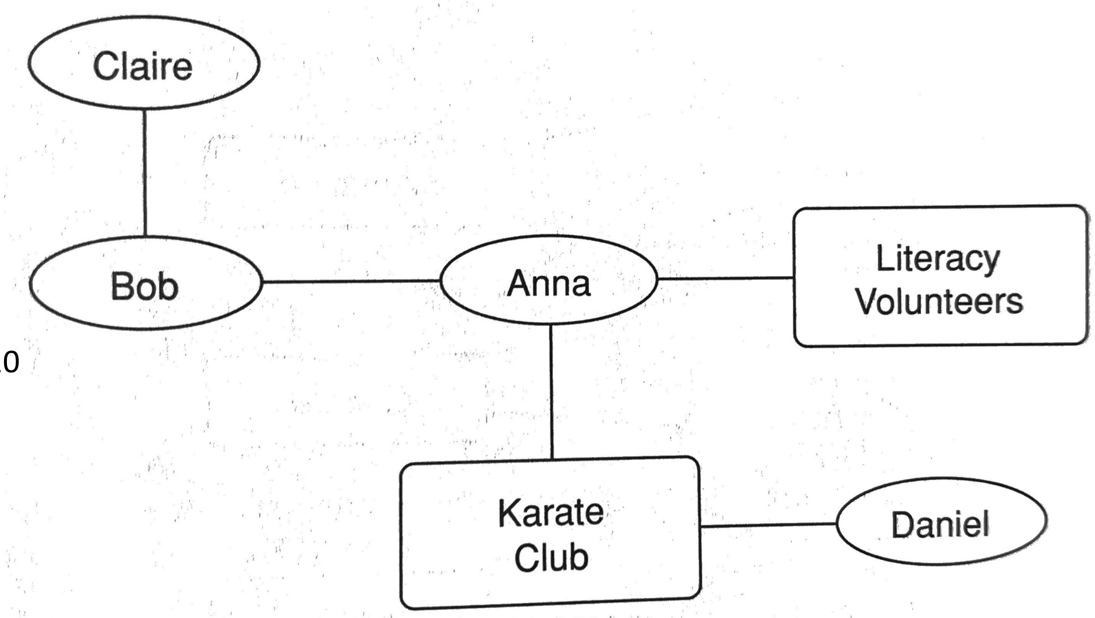
Easley and Kleinberg 2010

People on the left were on the board of directors of companies on the right.

What type of network is this? It is a bipartite affiliation network.

Having common affiliations will promote interactions of the “actors”

Closure is Influenced by Affiliation

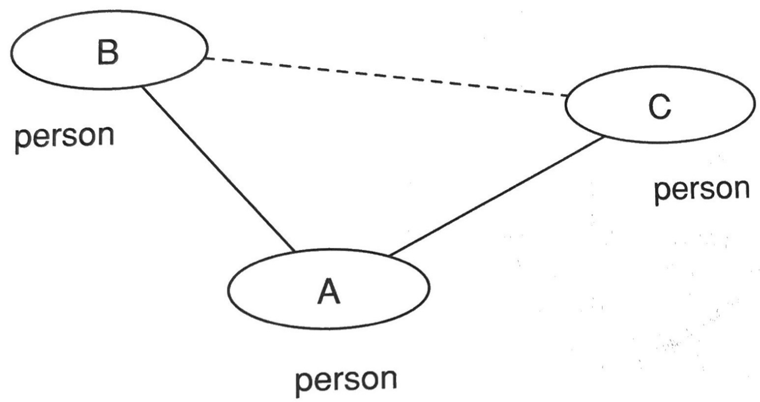


Easley and Kleinberg 2010

Simple example of a social-affiliation network. There are now edges between actors, as well as between actors and “foci” (something that brings people together)

Closure is Influenced by Affiliation

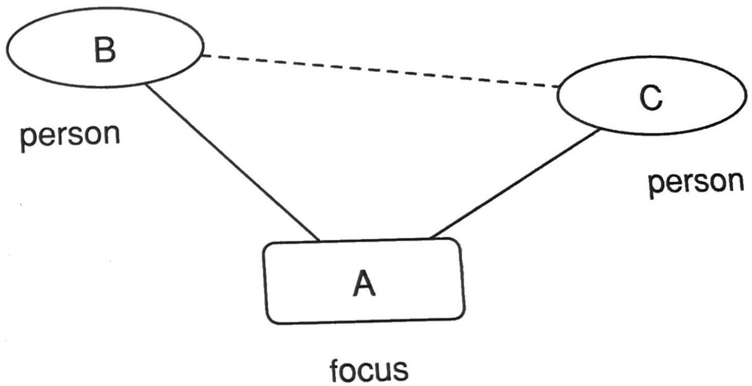
Easley and
Kleinberg 2010



Triadic closure

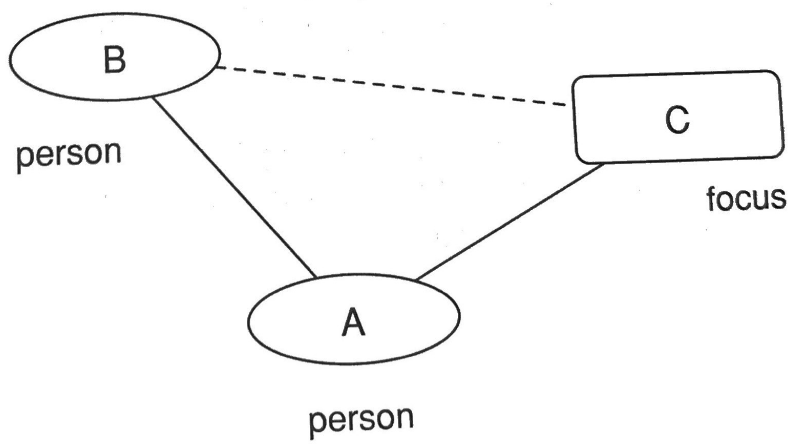
Closure is Influenced by Affiliation

Easley and
Kleinberg 2010



Focal closure

Closure is Influenced by Affiliation

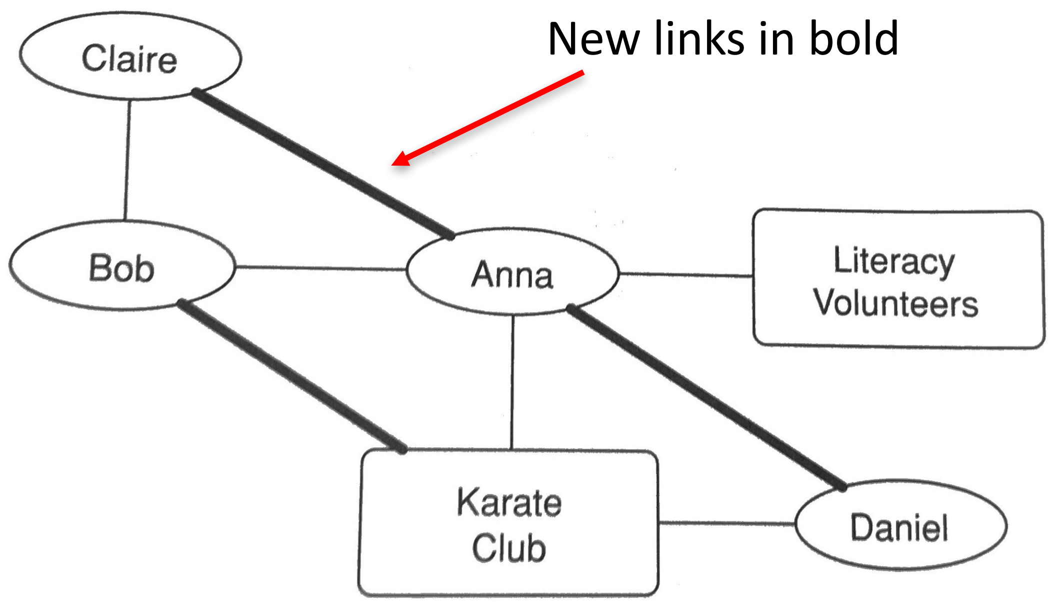


Easley and Kleinberg 2010

Membership closure

Closure is Influenced by Affiliation

Easley and
Kleinberg 2010



In real life, all of these connections can form social links.

Testing for Closure in a Real Social Network

The rationale for links in real social networks is rarely documented, and many people have trouble even remembering how they formed links with their friends and acquaintances. This type of information is stored for online interactions!

Strategy: Using online data,

(1) Take two snapshots of the network at different times

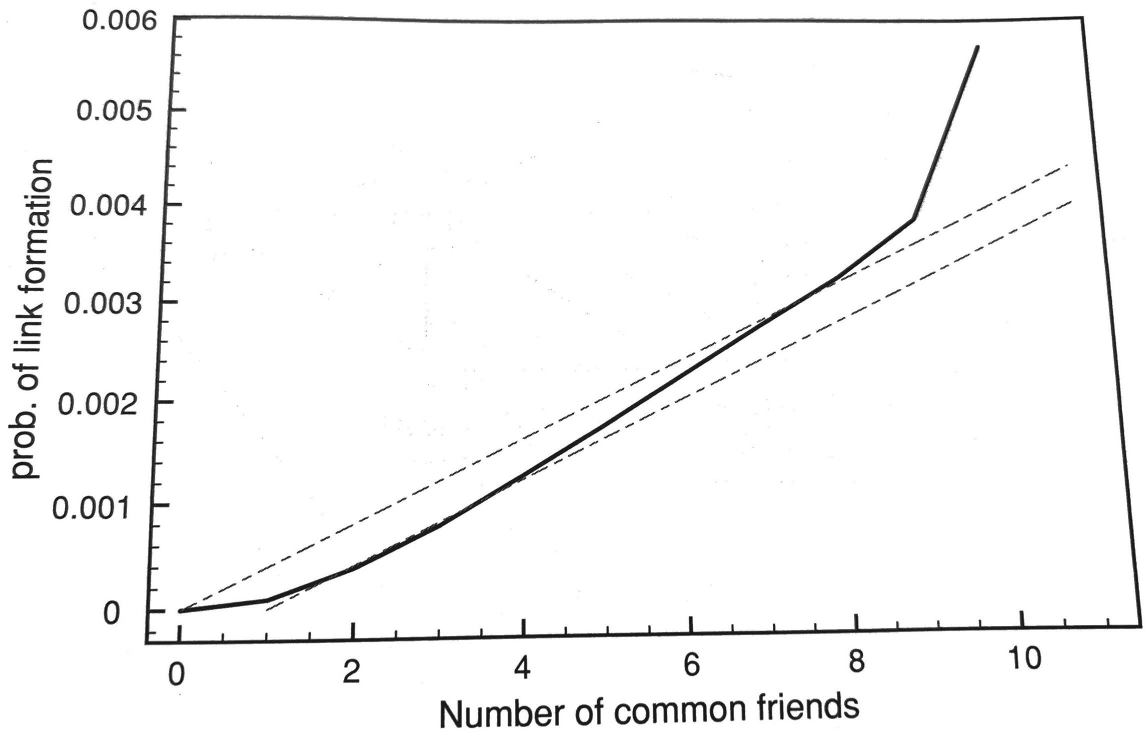
(2) For each *k*, identify all pairs of nodes who have exactly *k* friends in common in the first snapshot, but who are not directly connected

(3) Define *T(k)* to be the fraction of these pairs that have formed an edge by the time of the second snapshot. This approximates the probability that a link will form between two people with *k* common friends

(4) Plot *T(k)* as a function of *k* to illustrate the effect of common friends on the formation of links

Triadic Closure with E-mail Data

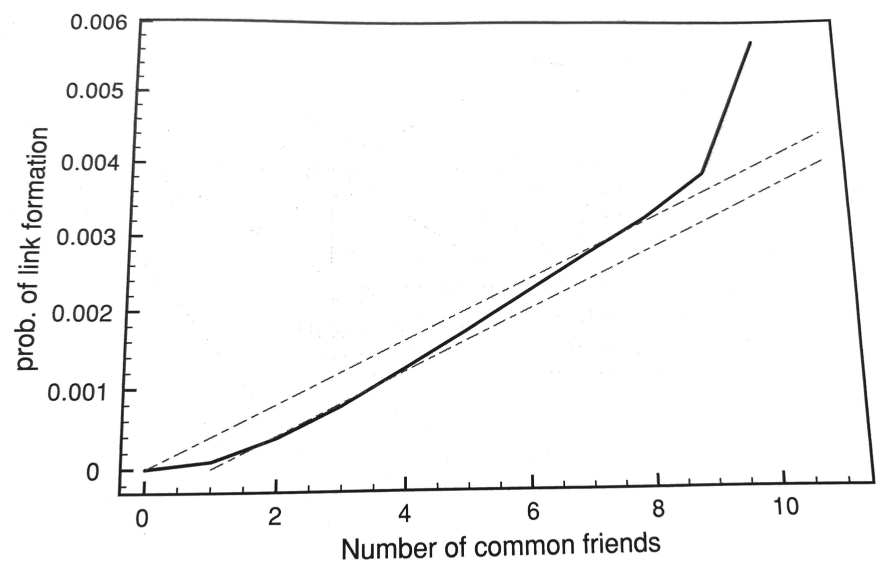
A study by Kossinets and Watts did this, using email communications among 22,000 students at a U.S. university over a one-year period. Two people were considered friends if one sent an email to the other during the preceding 60 days. They took multiple snapshots at different days throughout the year and averaged them to compute *T(k)*. The result is below.



Science, 311, (2006)

Solid curve is actual data

Triadic Closure with E-mail Data



Science, 311, (2006)

Let *p*=Prob[two people with a common friend forms a link]

Assume that each common friend gives this probability, independent of any other common friends.

Then probability of having *k* common friends and not forming a link is $(1-p)^k$

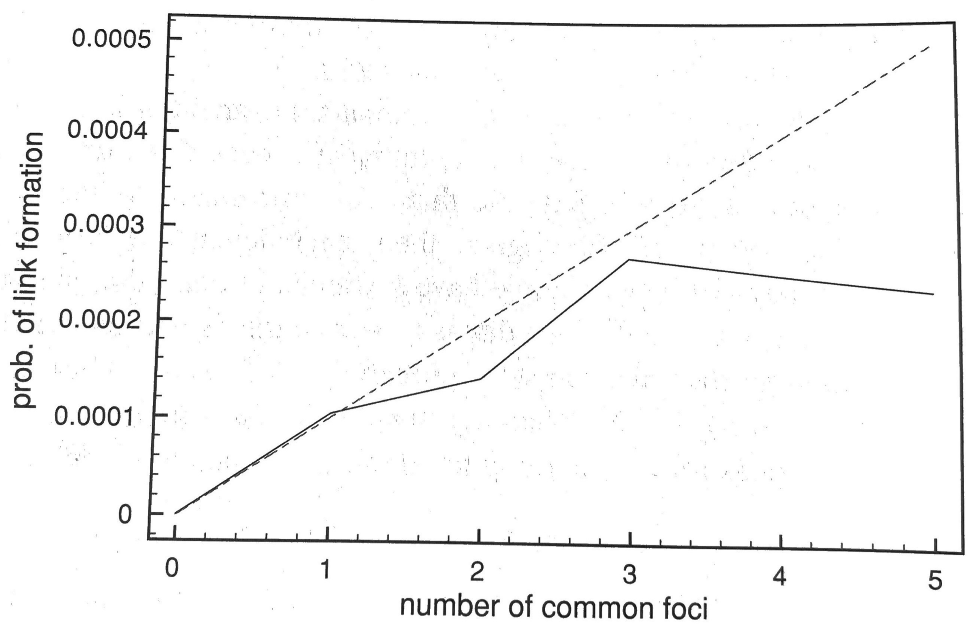
The probability that a link does form is then $T_{ind} = 1-(1-p)^k$. This is plotted as the upper dashed curve in the figure above.

The actual increase in link formation is greater than that from the model, so the probability of link formation with two common friends is more than expected from independent effects of the two common friends. There is a synergy.

Focal Closure with E-mail Data

In the study by Kossinets and Watts, they examined focal closure using classes as foci for the students (two students share a focus if they are in the same class).

Now *T(k)* is the probability they form a link (become email friends) if they have *k* classes together. Result is below.

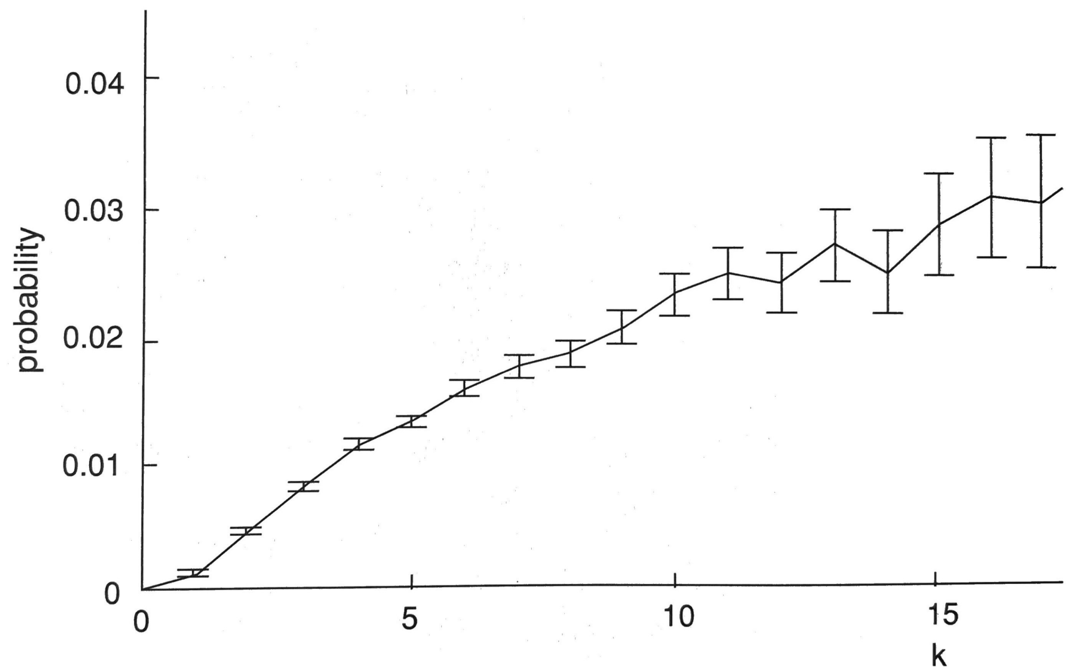


Science, 311, (2006)

While focal closure is evident for the first few values of *k*, it drops below what would be expected from the model that assumes independence. That is, there is a "diminishing returns" effect.

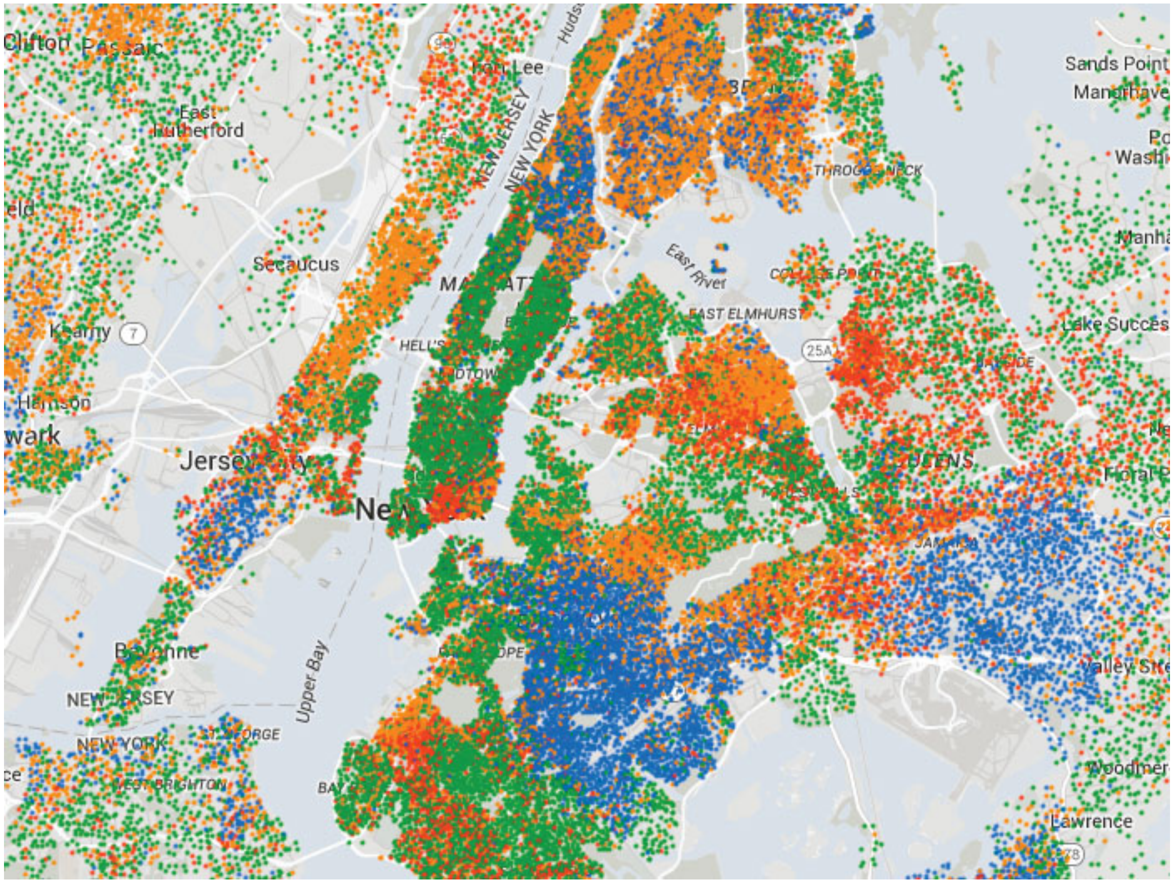
Membership Closure with Wikipedia Data

In a study on membership closure, a Wikipedia article is defined as a focus. A person is associated with that focus if he/she has edited it. If someone is friends with (i.e., has e-mail communication with) *k* editors of a Wikipedia article, what is the probability (*T(k)*) that that person becomes an editor of the article?



It is clear that membership closure occurs here, even for large *k*.

Spatial Segregation in Social Networks



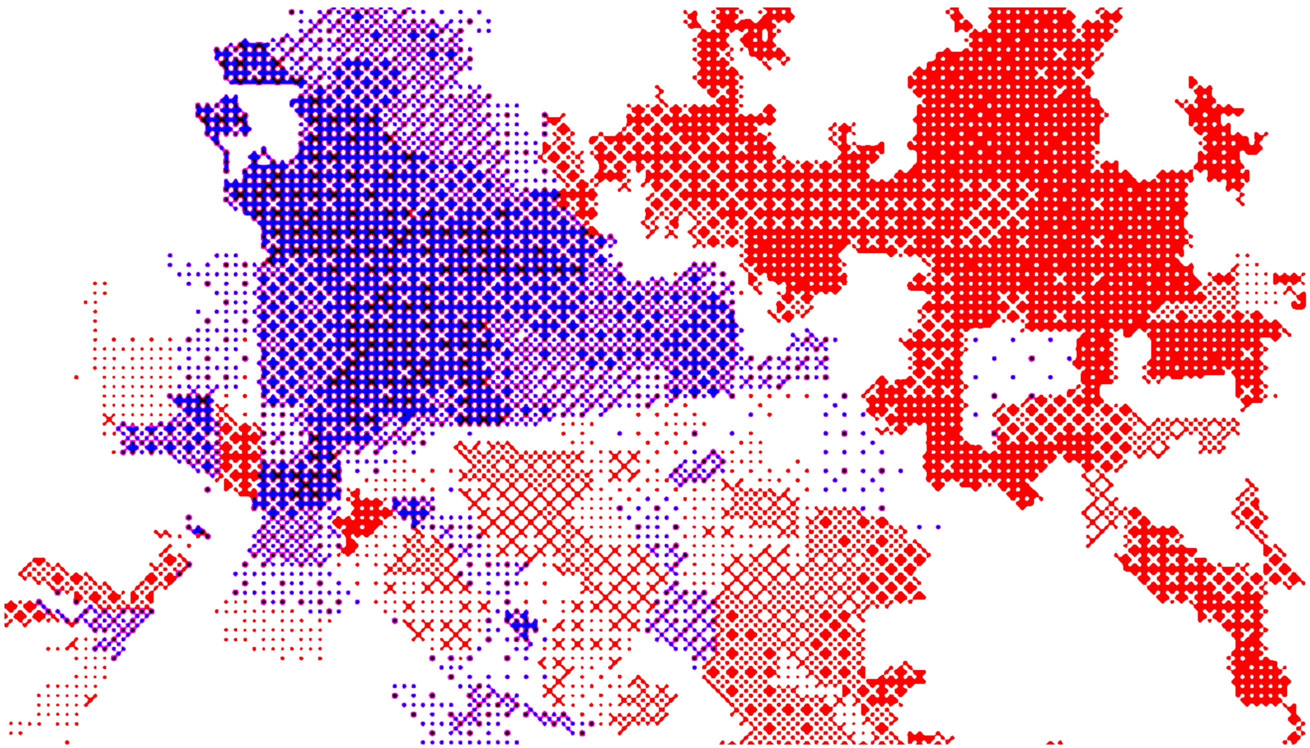
Segregation in Society

The concept of homophily states that people tend to form links with those that share common characteristics (race, ethnicity) or beliefs (political or religious). For example,

Most neighborhoods are currently filled with Biden signs, or filled with Trump signs. Few have a mix of the two.

Segregation in Society

Segregation in Democratic (blue) and Republican (red) households in a community.



Segregation in Society

Segregation in housing in Chicago. Blocks with lighter colors have the smallest percentages of African-Americans. Left=1940, Right=1960

Easley and Kleinberg 2010

Is Segregated Housing Intentional?

This question was addressed indirectly using computer simulations with the Schelling model (1972). This is a simple model that tests how small individual choices can lead to segregated housing.

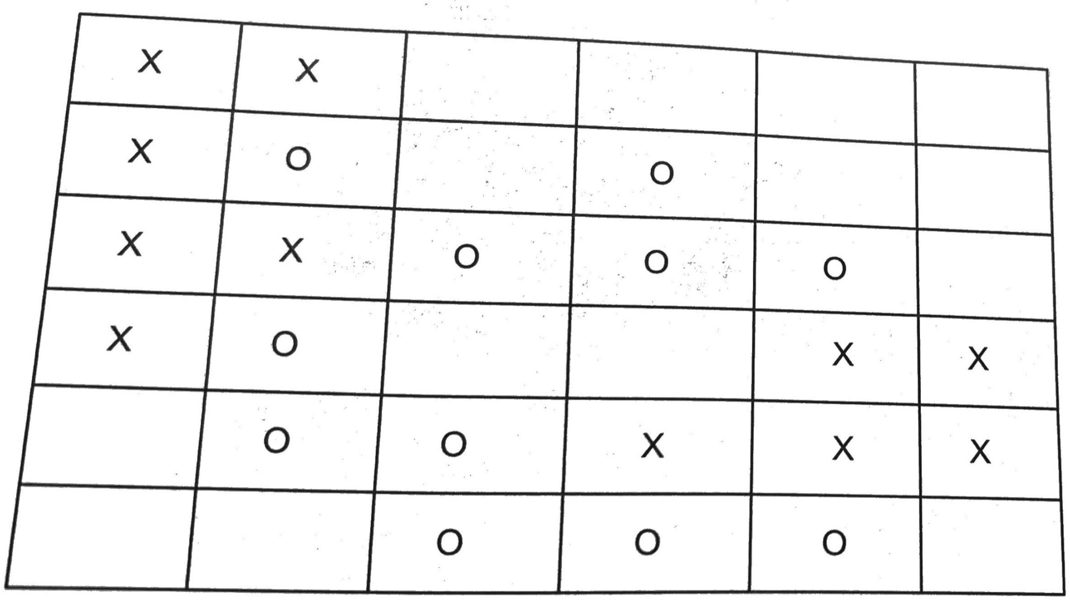
Assume a population of individuals ("agents") of type X or O. These agents live on a grid of cells representing the two-dimensional geography of a city.

"Neighbors" of a cell are those that touch it (up to 8 neighbors).

Some fraction of the cells are empty.

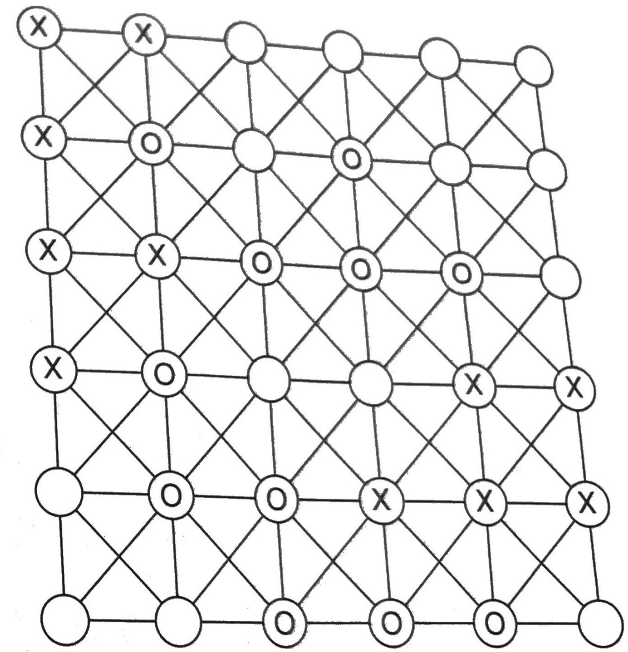
This grid can also be thought of as a network, with agents as nodes and physical neighbors as edges.

The Schelling Model



(a)

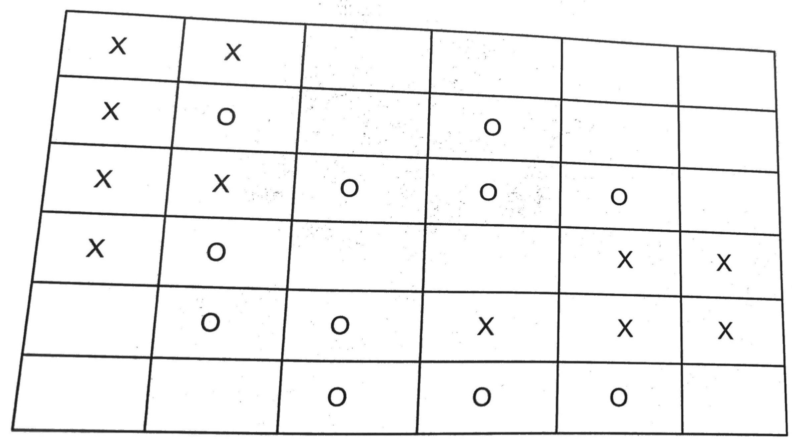
As a grid



(b)

As a network

The Schelling Model



(a)

Movement: An agent is ”unsatisfied” if fewer than T neighbors are of its type. In each round, each unsatisfied agent has the opportunity to move into a new location where it will be satisfied. The order in which this is done varies with different implementations.

The Schelling Model

X1*	X2*				
X3	O1*		O2		
X4	X5	O3	O4	O5*	
X6*	O6			X7	X8
	O7	O8	X9*	X10	X11
		O9	O10	O11*	

(a)

One round

X3	X6	O1	O2		
X4	X5	O3	O4		
	O6	X2	X1	X7	X8
O11	O7	O8	X9	X10	X11
	O5	O9	O10*		

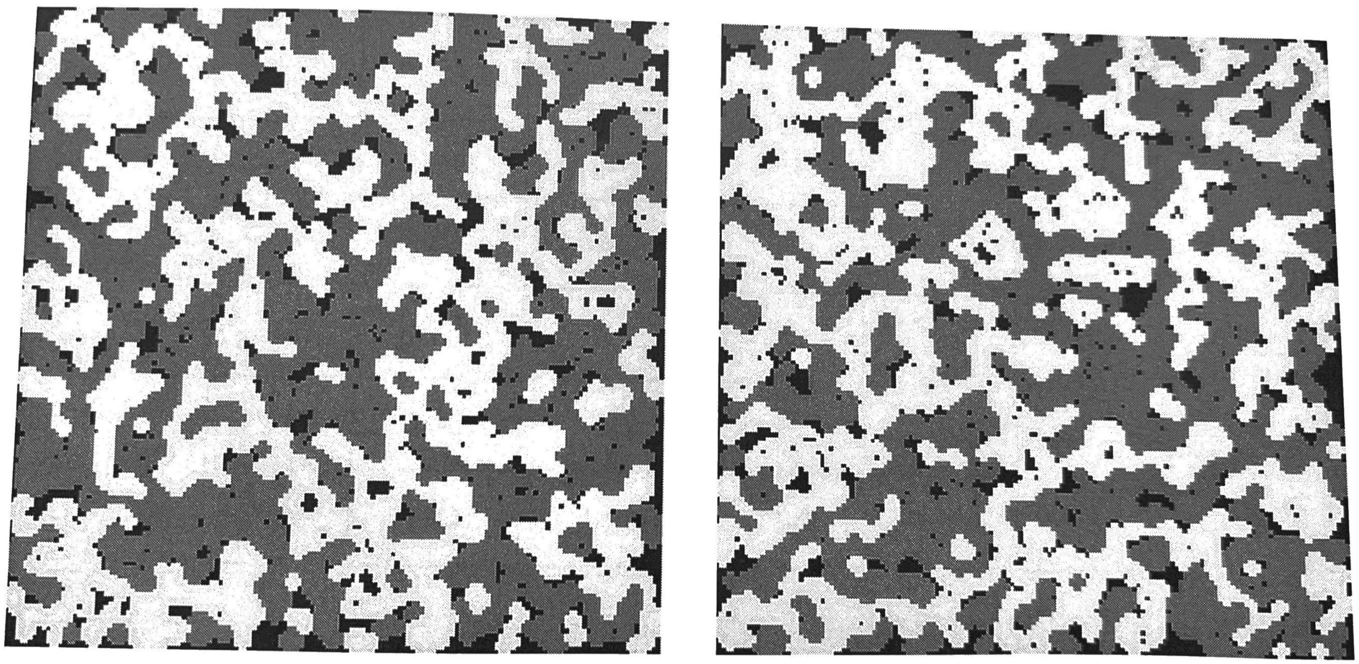
(b)

Unsatisfied agents (* superscript) moved and became satisfied. This left other agents unsatisfied. In this case, T=3. Agents labeled according to this starting location.

Computer Simulation with the Schelling Model

In this example, 10,000 agents are used and the grid is 150 X 150. The threshold for satisfaction is T=3. Start with random location of agents. Do this twice, each with a different initial random location of agents.

Easley and Kleinberg 2010



(a) (b)

Light dots=type O, gray dot=type X, black dot=empty

Patterns have formed, reflecting spatial segregation into types.

Computer Simulation with the Schelling Model

This did not have to happen! With T=3 it is possible to arrange agents into an integrated pattern:

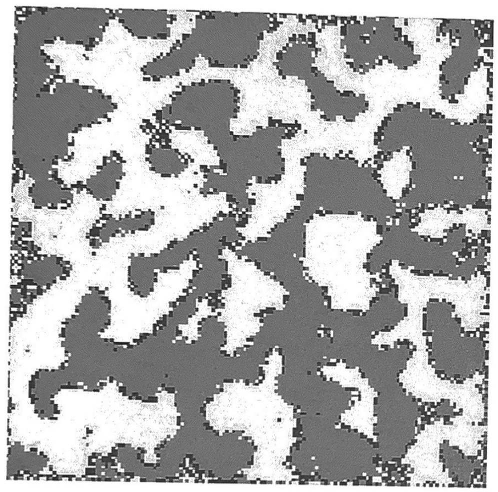
X	X	O	O	X	X
X	X	O	O	X	X
O	O	X	X	O	O
O	O	X	X	O	O
X	X	O	O	X	X
X	X	O	O	X	X

Everyone is satisfied, and there are no large clusters of agents of the same type. However, reaching this integrated distribution is very difficult from initial random placement of agents.

Computer Simulation with the Schelling Model

With threshold T=4 the segregation gets much worse.

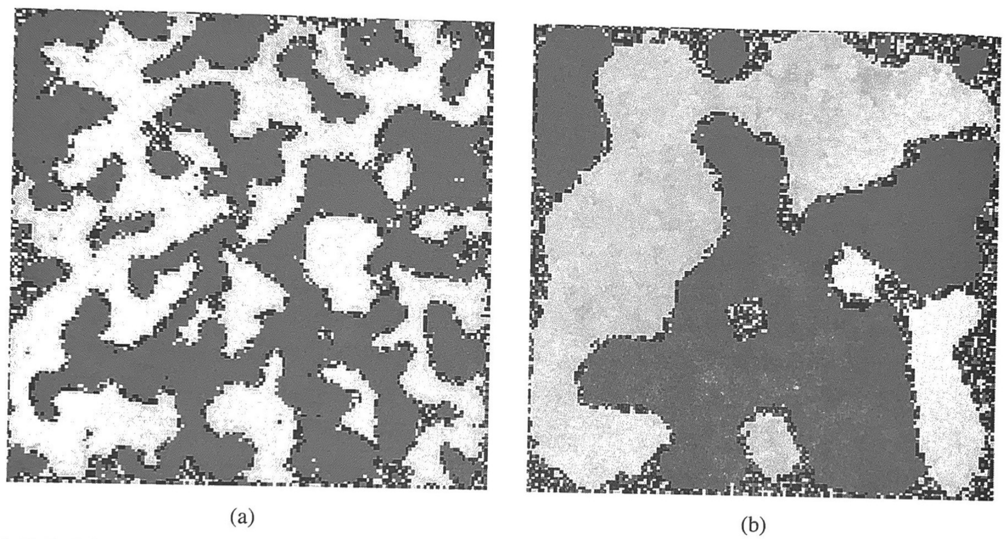
20 rounds



Computer Simulation with the Schelling Model

With threshold T=4 the segregation gets much worse.

20 rounds

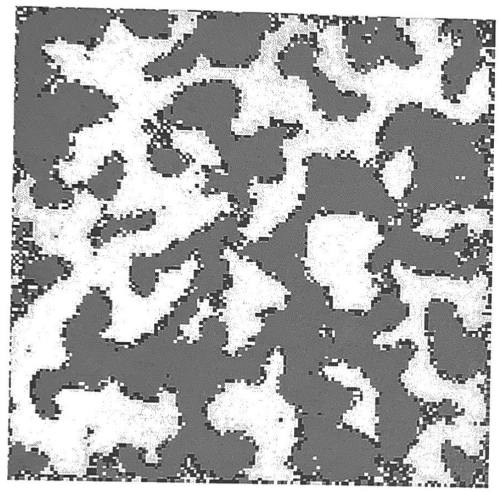


150 rounds

Computer Simulation with the Schelling Model

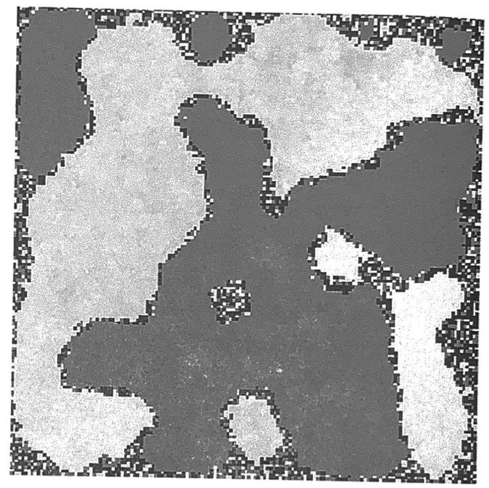
With threshold T=4 the segregation gets much worse.

20 rounds



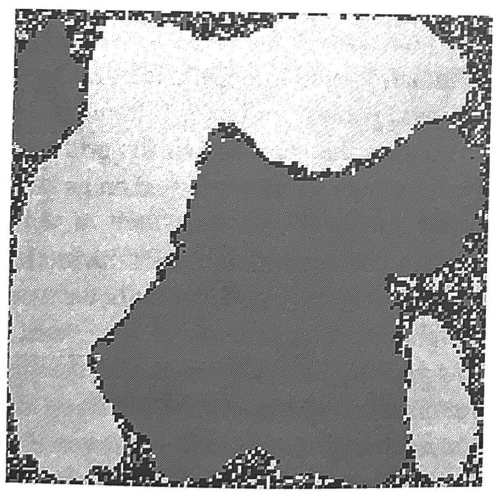
(a)

150 rounds



(b)

350 rounds



(c)

Computer Simulation with the Schelling Model

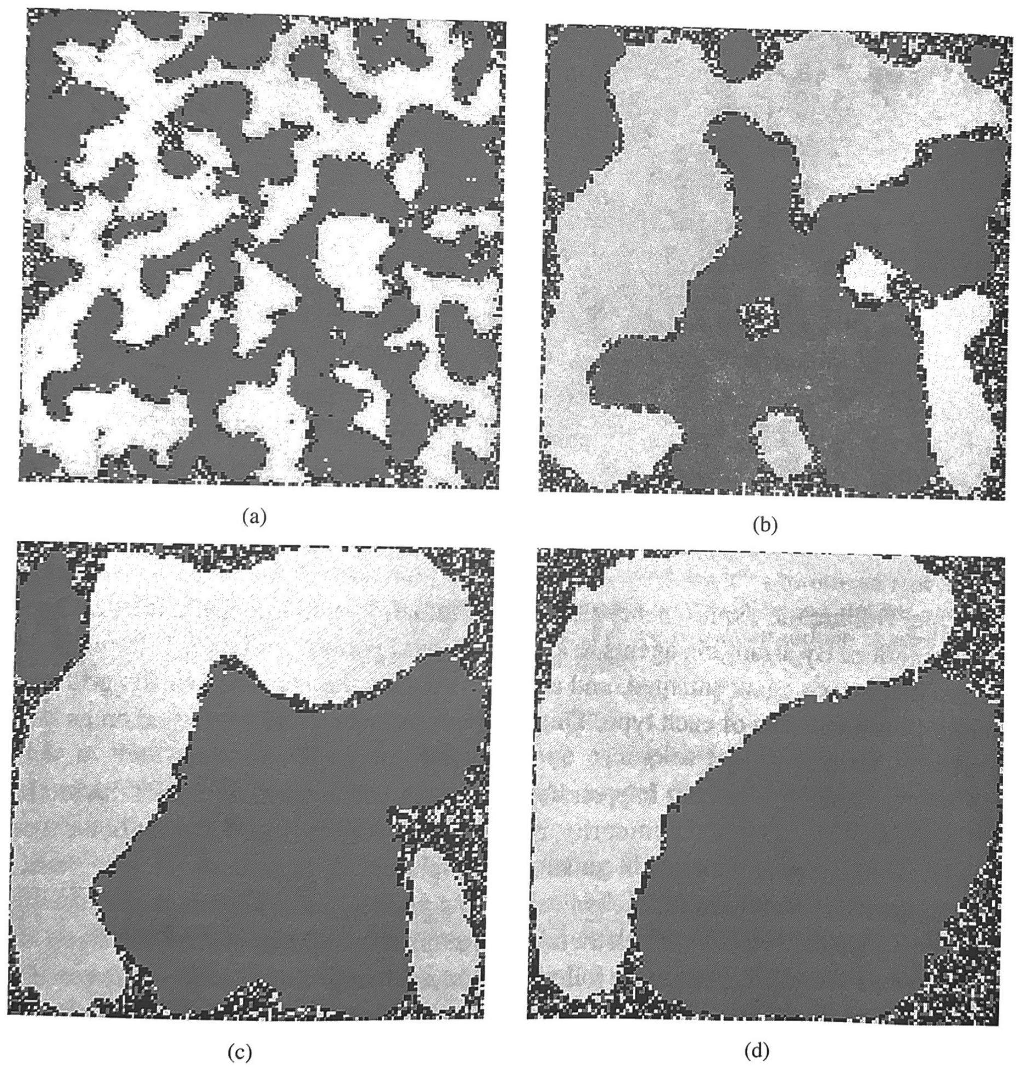
With threshold T=4 the segregation gets much worse.

20 rounds

150 rounds

350 rounds

800 rounds



Easley and
Kleinberg 2010

Important Observations from the Schelling Model

(1) With T=3 or 4, none of the agents minded being in the minority. For example, with T=3, an agent would be satisfied with having 3 neighbors of its type and 5 neighbors of the other type.

(2) The agents **did not plan to move into segregated clusters**, these just emerged over time in response to many individual moves to satisfy local preferences. It is an example of an emergent property of the system.

(3) Computer simulations with this model show that the underpinnings of segregation are present in a system where individuals simply want to avoid being in too extreme a minority in their own local area.

The End